Chemometric Methods for Prediction of Uncertainties and Spectral Validation of Rank Deficient Mechanisms in Kinetic Hard-modelling of Spectroscopic Data

Julien Billeter

Diss. ETH No. 18311 (2009)

Diss. ETH No. 18311

Chemometric Methods for Prediction of Uncertainties and Spectral Validation of Rank Deficient Mechanisms in Kinetic Hard-modelling of Spectroscopic Data

A dissertation submitted to

ETH ZURICH

for the degree of
Doctor of Sciences

presented by

Julien Léo Billeter

Ing. Chim. Dipl. EPF Lausanne
born on June 24, 1978
citizen of Männedorf - ZH

accepted on the recommendation of

Prof. Dr. Konrad Hungerbühler, examiner
Prof. Dr. Manfred Morari, co-examiner
Dr. Yorck-Michael Neuhold, co-examiner

Zurich 2009

"Toute science crée une nouvelle ignorance"

"Any science leads to a new ignorance"

"Jede Wissenschaft führt zu einer neuen Unkenntnis"

Henri Michaux
French poet
1899 – 1984

CONTENT

Preface v
Abstract vii
Zusammenfassung ix
Résumé xi
Notation xiii
Acronyms xix

1 Introduction 1

1.1 Overview . . . 1
1.2 Motivation . . . 1
1.3 Objectives . . . 3
1.4 Structure . . . 4
1.5 Appendix . . . 6
1.5.1 Condensation of carbonyls into hydrazone derivatives . . . 6
1.5.2 Condensation of carbonyls into enamine derivatives . . . 7
1.6 References . . . 9

2 Instrumental and chemometric tools in chemical kinetics 11

2.1 Instrumental tools . . . 11
2.1.1 Calorimetry . . . 11
2.1.2 Spectroscopy . . . 14
2.1.2.1 mid-IR spectroscopy . . . 18
2.1.2.2 UV-vis spectroscopy . . . 19
2.2 Chemometric tools for kinetic analysis . . . 20
2.2.1 Kinetic hard-modelling . . . 20
2.2.1.1 Modelling **C** or **A** . . . 20
2.2.1.2 Direct kinetic hard-modelling . . . 21
2.2.1.3 Indirect (inverse) kinetic hard-modelling . . . 22
2.2.1.4 Implicit calibration . . . 22
2.2.1.5 Explicit calibration . . . 24
2.2.1.6 Non-linear optimisation algorithms . . . 24
2.2.2 Kinetic soft-modelling . . . 26
2.2.2.1 Factor analysis . . . 26
2.2.2.2 Target factor analysis . . . 28
2.2.2.3 Evolving factor analysis . . . 30
2.2.2.4 Alternating least squares . . . 31
2.2.3 Hybrid methods . . . 31
2.3 Appendix . . . 32
2.3.1 Data structure and measurement order . . . 32
2.3.2 Bilinearity, trilinearity and n-linearity . . . 32
2.3.3 Definition of linear and non-linear parameters . . . 33
2.3.4 Left and right pseudo-inverse of a matrix . . . 34
2.4 References . . . 36

3 Uncertainties and error propagation in kinetic hard-modelling of spectroscopic data 39

3.1 Abstract . . . 39
3.2 Introduction . . . 40
3.3 Theoretical considerations . . . 42
3.4 Simulations and experiments . . . 47
3.4.1 Simulations . . . 47
3.4.2 Experiments . . . 49
3.4.2.1 Sample preparation . . . 50
3.4.2.2 Instrumentation . . . 51
3.5 Results and discussion . . . 52
3.5.1 Simulated data . . . 52
3.5.1.1 Second order model . . . 52
3.5.1.2 Third order model . . . 56
3.5.2 Experimental data . . . 57
3.6 Conclusion . . . 63
3.7 Appendix . . . 65
3.7.1 Error estimation in the initial concentrations due to sample preparation . . . 65
3.7.2 Analytical derivatives of a second order rate constant with respect to the initial concentrations . . . 66
3.8 References . . . 68

4 Systematic prediction of linear dependencies in the concentration profiles and implications on the kinetic hard-modelling of spectroscopic data 71

4.1 Abstract . . . 71
4.2 Introduction . . . 72
4.3 Theory . . . 75
4.3.1 Rank deficiencies in the concentration matrix **C** . . . 79
4.3.1.1 Kernel of **C** . . . 79
4.3.1.2 A pseudo-equivalent matrix for **C** under semi-batch conditions . . . 81
4.3.2 A systematic method for selecting the appropriate strategy to treat rank deficiencies in **C** . . . 82
4.3.2.1 Prediction of the uncoloured species to include in Strategy (1) . . . 85
4.3.2.2 Prediction of the known pure spectra to provide in Strategy (2) . . . 88
4.3.2.3 Prediction of the species to dose in Strategy (3) . . . 89
4.3.2.4 Prediction of the initial concentrations to vary in Strategy (4) . . . 90
4.4 Examples . . . 91
4.4.1 Example 1: a second order reaction . . . 92
4.4.1.1 Rank augmentation by dosing (Strategy 3) . . . 95
4.4.2 Example 2: a mechanism with two parallel consecutive reactions . . . 95
4.4.2.1 Model reduction by defining uncoloured species (Strategy 1) . . . 97
4.4.2.2 Model reduction by providing a pure spectrum (Strategy 2) . . . 100
4.4.3 Example 3: a mechanism with three parallel consecutive reactions . . . 100
4.4.3.1 Model reduction by defining uncoloured species (Strategy 1) . . . 104
4.4.3.2 Model reduction by providing known pure spectra (Strategy 2) . . . 106
4.4.3.3 Rank augmentation by dosing (Strategy 3) . . . 107
4.4.3.4 Rank augmentation by varying initial concentrations (Strategy 4) . . . 109
4.5 Conclusion . . . 111

4.6 Appendix . . . 113
4.6.1 Reduction of time variant Equation 4.15 to time invariant Equation 4.16 . . . 113
4.6.1.1 Elimination of diagonal matrices **D** and (**I** – **D**) . . . 113
4.6.1.2 Substitution of time variant $\mathbf{C_0}$ and $\mathbf{C_{dos}}$ by time invariant $\mathbf{c_0}$ and $\mathbf{C_{in}}$ . . . 113
4.6.1.3 Pseudo-equivalence between the reaction extent and its derivative . . . 114
4.6.1.4 Pseudo-equivalence between the reaction extent and the rate law . . . 115
4.6.1.5 Simplification of the time variant rate law into a time invariant rate law . 115
4.6.2 Matlab example for the setup of $\mathbf{\Omega}$ applied to the kinetic scheme of Figure 4.1 and the calculation of its kernel under batch conditions . . . 117
4.6.3 Comparison of ker $\mathbf{\Omega}$ with ker **C** for the kinetic scheme of Equation 4.39 under batch conditions . . . 117
4.6.4 Kernel of $\mathbf{\Omega}$ when strategies (1) to (4) are applied to the kinetic scheme of Eq. 4.39 118
4.6.5 Prediction of linear dependencies in the reaction extents and implications on the kinetic hard-modelling of calorimetric data . . . 119
4.6.5.1 Application of Strategy (1) – defining non-contributing reactions . . . 120
4.6.5.2 Application of Strategy (2) – including known enthalpies into the analysis 121
4.7 References . . . 123

5 Kinetic hard-modelling and spectral validation of rank deficient spectroscopic data: a case study 125

5.1 Abstract . . . 125
5.2 Introduction . . . 126
5.3 Theoretical background . . . 127
5.3.1 Strategy (1) – defining uncoloured species . . . 130
5.3.2 Strategy (2) – including known spectra into the analysis . . . 130
5.3.3 Strategy (3) – dosing one or more species . . . 130
5.3.4 Strategy (4) – performing a second order global analysis . . . 131
5.4 Experiments . . . 131
5.4.1 Instruments . . . 132
5.4.2 Sample preparation . . . 132
5.4.3 Independently measured pure component spectra . . . 132
5.4.4 Experimental conditions (a) – dosing *Aa* . . . 133
5.4.5 Experimental conditions (b) – dosing *B* (1.986 mol·L^{-1}) . . . 133
5.4.6 Experimental conditions (c) – sequential dosing of *B* and *Aa* . . . 135
5.4.7 Experimental conditions (d) – dosing *P* . . . 135
5.4.8 Experimental conditions (e) – dosing *B* (1.192 mol·L^{-1}) . . . 135
5.5 Results and discussion . . . 136
5.5.1 Analytical solution for $\mathbf{\Omega}$ . . . 136
5.5.2 Independently measured pure component spectra . . . 138
5.5.3 Application of Strategy (1) – defining uncoloured species . . . 140
5.5.4 Application of Strategy (2) – including known spectra into the analysis . . . 144
5.5.5 Application of Strategy (3) – dosing one or more species . . . 147
5.5.6 Application of Strategy (4) – performing a second order global analysis . . . 148
5.6 Conclusion . . . 151
5.7 Appendix . . . 153
5.7.1 Matlab code for Section 5.5.1 . . . 153
5.7.2 Matlab code for Section 5.5.3 . . . 153
5.7.3 Matlab code for Section 5.5.4 . . . 153
5.7.4 Matlab code for Section 5.5.5 . . . 153
5.7.5 Matlab code for Section 5.5.6 . . . 154
5.7.6 Calculation of ker $\mathbf{\Omega}$ under batch conditions using mass balance equations . 154
5.7.7 Calculation of ker $\mathbf{\Omega}$ under semi-batch conditions using mass balance equations . . 155
5.8 References . . . 158

6 Conclusion and outlook **159**

6.1 Conclusion 159
6.2 Outlook 161

Citation **163**

Curriculum vitae **165**

PREFACE

The present thesis is the result of a research performed in the group of Safety and Environmental Technology at the Swiss Federal Institute of Technology (ETH) in Zurich between May 2005 and May 2009. The core of this thesis is based on three original papers published prior to the release of this document.

I would like to thank Prof. Konrad Hungerbühler, who gave me the opportunity of doing my PhD in his research group and funded this doctoral project. I also thank Prof. Manfred Morari for his assistance as co-examiner.

My deep gratitude goes to my direct supervisor, Dr. Yorck-Michael (Bobby) Neuhold, leader of the subgroup of Reaction Analysis, who greatly contributed to this research by his vast experience in chemometrics and spectroscopy. I particularly appreciated the help he provided during the writing process and the fruitful discussions we had together all along this work. I am also particularly grateful to my previous supervisor, Dr. Graeme Puxty, who introduced me in the field of kinetic hard-modelling and spectral fitting.

I would like to acknowledge the help of Dr. Gilles Richner for his help maintaining the reaction calorimeter, and the contributions of Dr. Stavros Papadokonstantakis and Dr. Levente Simon for their insights into chemical engineering. My thanks are extended to my colleagues in the subgroup of Reaction Analysis, Stefano Gianoli, Sébastien Cap and Tamás Godány, and to all other members in the group of Safety and Environmental Technology.

Finally, my greatest appreciation is reserved to my mother for her love, support and unfailing faith in me.

Julien Billeter

Zurich, March 2009

ABSTRACT

This dissertation contributes to the development of novel chemometric methods for the kinetic modelling of multivariate and isothermal spectroscopic data using hard kinetic models, i.e. based on first-principles models. More specifically, it has the aims to facilitate the analysis of rank deficient spectroscopic data, to enable the validation of rank deficient kinetic models, and to improve the prediction of uncertainties in the optimised associated kinetic parameters, e.g. the rate constants.

The development of these chemometric methods is driven by the needs of chemical and pharmaceutical industry for efficient kinetic models in order to design and optimise the manufacturing conditions. These optimum conditions contribute to enhance product quality, reduce costs, and improve safety, health and environmental aspects of industrial processes.

The two main contributions of this doctoral thesis are the extension of the calculated errors in the optimised non-linear kinetic parameters to include experimental uncertainties, and the systematic identification of linear dependencies in the kinetic concentration profiles allowing the spectral validation of rank deficient mechanisms. Simulated and experimental data are used to show the impact of both contributions on the design of experiments to be analysed by kinetic hard-modelling. In-situ experimental data, recorded in the mid-IR and UV-vis range, are obtained by following the reaction of benzophenone and phenylhydrazine to the corresponding hydrazone under acetic acid catalysis in the well-defined environment of an isothermal reaction calorimeter (CRC.v4).

Uncertainties and error propagation in kinetic hard-modelling

Gradient-based algorithms used for non-linear optimisation of kinetic parameters allow estimating the uncertainty in the fitted parameters. However, this estimation is only based on the measured spectroscopic data and is systematically underestimated compared to the uncertainty resulting from multiple repetitions of the experiment under the same conditions. In this thesis, a method is presented for the rigorous propagation of uncertainties in the experimental conditions (e.g. initial concentrations, dosing rates) into the errors of the rate constants fitted by multivariate kinetic hard-modelling of spectroscopic data. Moreover, this method also allows predicting the experimental conditions, for which uncertainties in the rate constants are minimal.

Systematic prediction of linear dependencies in the concentration profiles and spectral validation of rank deficient mechanisms

Linear dependencies in concentration profiles due to rank deficiencies in kinetic mechanisms represent a mathematical dilemma when spectroscopic data are directly fitted and implicitly used, together with the modelled concentration profiles, as a calibration for the fitting of absorptivity spectra. Thus, a novel method has been developed for the systematic identification of the linear dependencies in the concentration matrix and for the prediction of the minimum requirements to enable the optimisation of kinetic parameters and absorptivity spectra associated with the corresponding rank deficient model. This method is based on a novel time invariant matrix covering potential rank deficiency due to stoichiometry and rate laws. It allows predicting the resolution in the fitted absorptivity spectra in case of model reduction or rank augmentation, the two major strategies to deal with linear dependencies in the concentration profiles.

ZUSAMMENFASSUNG

Diese Dissertation trägt zur Entwicklung neuer chemometrischer Methoden bei, die zur kinetischen Modellierung von isothermen multivariaten spectroskopischen Daten mit Hilfe harter kinetischer Modelle, basierend auf physikalisch-chemischen Prinzipien, verwendet werden können. Ziele dieser Arbeit sind die Erleichterung der Analyse rangdefizienter spectroskopischer Daten, die Validierung rangdefizienter kinetischer Modelle, und die verbesserte Abschätzung der Unsicherheiten in den optimierten kinetischen Parametern, z. B. den Geschwindigkeitskonstanten.

Die treibende Kraft hinter der Entwicklung dieser chemometrischer Methoden ist der erhöhte Bedarf der chemischen und pharmazeutischen Industrie nach effizienten kinetischen Modellen zum Design und zur Optimierung von Reaktionsbedingungen. Optimierte Reaktionsbedingungen in industriellen Verfahren tragen zur derer Qualitätsverbesserung und Kostenreduktion bei, und wirken sich positiv auf Sicherheit, Gesundheit und Umwelt aus.

Die beiden Hauptbeiträge dieser Doktorarbeit beinhalten, experimentelle Unsicherheiten in die Fehler der optimierten kinetischen Parameter einzubeziehen, und lineare Abhängigkeiten der Konzentrationsprofile systematisch zu identifizieren, um die spektrale Validierung rangdefizienter kinetischer Mechanismen zu ermöglichen. Mit Hilfe simulierter und experimenteller kinetischer Daten werden deren Einflüsse auf das experimentelle Design aufgezeigt. Experimentelle in-situ Daten für die Essigsäure katalysierte Reaktion zwischen Benzophenon und Phenylhydrazin unter Bildung des Hydrazons wurden sowohl im mittleren Infrarot- als auch im Ultraviolett-/sichtbaren Bereich in der sehr gut definierten Umgebung eines isothermen Reaktionskalorimeters (CRC.v4) aufgenommen.

Unsicherheiten und Fehlerfortpflanzung bei harter kinetischer Modellierung

Gradienten basierende Algorithmen zur nicht-linearen kinetischen Parameteroptimierung erlauben eine direkte Abschätzung von Unsicherheiten optimierter Parameter. Allerdings stützt sich diese Abschätzung bisher lediglich auf die gemessenen spektroskopischen Daten, und wird systematisch unterschätzt im Vergleich zu den Unsicherheiten, die man nach mehrfacher Wiederholung der Experimente unter denselben Bedingungen erhält. In dieser Doktorarbeit wird eine Methode vorgestellt, die eine robuste Fehlerfortpflanzung der experimentellen Bedingungen (z.B. der Anfangskonzentrationen oder Dosiergeschwindigkeit) auf die Unsicherheiten der optimierten Geschwindigkeitskonstanten ermöglicht. Zudem erlaubt diese Methode, diejenigen experimentellen Bedingungen zu bestimmen, welche zu möglichst geringen Unsicherheiten in den Geschwindigkeitskonstanten führen.

Systematische Vorhersage linearer Abhängigkeiten der Konzentrationsprofile und spektrale Validierung rangdefizienter Mechanismen

Lineare Abhängigkeiten von Konzentrationsprofilen, die sich aus rangdefizienten kinetischen Modellen ergeben, führen bei direkter impliziter Optimierung spektroskopischer Daten zu einem mathematischen Dilemma, da die Messdaten selbst zusammen mit den modellierten Konzentrationsprofilen als implizite Kalibrierung für die Absorptionsspektren benutzt werden. Auf diesem Hintergrund wurde eine neue Methode zur systematischen Identifizierung linearer Abhängigkeiten der Konzentrationsprofile entwickelt, welche zudem eine Vorhersage derjenigen Bedingungen ermöglicht, die eine Optimierung der kinetischen Parameter und der Absorptionsspektren für ein rangdefizientes kinetisches Modell erlaubt. Die Methode beruht auf der Bestimmung einer neuen zeit-invarianten Matrix, welche lineare Abhängigkeiten sowohl in Stoichiometrie als auch im Geschwindigkeitsgesetz berücksichtigt. Dies erlaubt die Vorhersage der möglichen Auflösung der Absorptionsspektren der einzelnen Spezies, wenn die beiden wichtigsten Strategien – Modelreduktion und Rangerhöhung – zur Behandlung linearer Abhängigkeiten von Konzentrationsprofilen angewandt werden.

RÉSUMÉ

Cette dissertation traite du développement de nouvelles méthodes chimiométriques applicables à la modélisation cinétique déterministe, c'est-à-dire reposant sur des principes physico-chimiques fondamentaux, de données multivariées et isothermes. Plus spécifiquement, ce travail a pour buts de faciliter l'analyse des données spectroscopiques à rang déficient, de permettre la validation des modèles cinétiques à rang déficient, et d'améliorer la prédiction des incertitudes associées aux paramètres cinétiques à optimiser, comme par exemple dans le cas de l'optimisation des constantes de vitesse.

Le développement de ces méthodes cinétiques est prescrit par les besoins de l'industrie chimique et pharmaceutique à l'égard de modèles cinétiques efficaces pouvant être utilisés pour le design et l'optimisation des conditions de production. Ces conditions optimales contribuent à l'augmentation de la qualité des produits, et à l'amélioration des procédés industriels quant à leurs conséquences sur la sécurité, la santé au travail et l'environnement.

Les deux principales contributions de cette thèse de doctorat sont l'extension du calcul des erreurs sur les paramètres cinétiques non-linéaires aux incertitudes expérimentales, et l'identification systématique des dépendances linéaires des profils de concentration. Cette identification permettant ensuite la validation spectrale des mécanismes cinétiques déficients de par leur rang. Des données tant simulées que mesurées sont utilisées dans cette dissertation pour démontrer l'impact de ces deux contributions sur le design des expériences analysées par modélisation cinétique déterministe. Les données expérimentales couvrant le domaine spectral de l'infrarouge moyen, de l'ultraviolet et de la lumière visible ont été obtenues en suivant in-situ la réaction de la benzophenone avec la phenylhydrazine formant, par catalyse avec l'acide acétique, l'hydrazone correspondant, cette réaction étant menée dans l'environnement contrôlé d'un calorimètre de réaction isotherme (CRC.v4).

Incertitudes et propagation des erreurs en modélisation cinétique déterministe

Les algorithmes basés sur le calcul du gradient sont fréquemment utilisés pour l'optimisation non-linéaire des paramètres cinétiques, car ils permettent l'estimation de l'incertitude sur les paramètres optimisés. Cependant, les incertitudes ainsi obtenues tiennent uniquement compte de l'erreur sur les données mesurées, ce qui conduit à les sous-estimer systématiquement en comparaison avec les incertitudes calculées en répétant les expériences dans des conditions identiques. Dans cette thèse, une méthode est présentée permettant la propagation rigoureuse des incertitudes liées aux conditions expérimentales (par exemple liées aux concentrations initiales ou à la vitesse de dosage) sur les erreurs des constantes de vitesse optimisées durant la modélisation cinétique déterministe de données spectroscopiques multivariées. De plus, cette méthode permet la prédiction des conditions expérimentales à employer pour minimiser les incertitudes sur les constantes de vitesse.

Prédiction systématique des dépendances linéaires des profils de concentration et validation spectrale des mécanismes déficients en rang

Les dépendances linéaires dans les profils de concentration des mécanismes cinétiques déficients en rang représentent un dilemme mathématique pour l'optimisation directe et implicite des données spectroscopiques, puisque les données elles-mêmes ainsi que les profils de concentration sont utilisés pour la calibration implicite des spectres d'absorptivité. C'est pourquoi une nouvelle méthode a été développée pour l'identification systématique des dépendances linéaires dans la matrice de concentration et pour la prédiction des informations minimales à fournir pour rendre possible l'optimisation des paramètres cinétiques et des spectres d'absorptivité associés à un modèle déficient en rang. Cette méthode est basée sur une matrice invariable en temps couvrant les déficiences de rang dus à la stœchiométrie et aux lois de vitesse. Elle permet de prédire le niveau de résolution attendu dans les spectres d'absorptivité calculés lorsque les deux stratégies les plus répandues pour le traitement des dépendances linéaires des profils de concentration – la réduction du modèle ou l'augmentation de rang – sont utilisées.

NOTATION

Convention

As a convention for the notation, multidimensional arrays (or tensors) are written in boldface capitals with a lower bar (e.g. $\underline{\mathbf{R}}$), matrices in boldface capitals (e.g. **R**), vectors in boldface lowercases (**r**) and scalars in italics (*r*). For indices, lowercase characters are used. Elements of an array $\underline{\mathbf{R}}$ are denoted as $r_{i,j,k}$, elements of a matrix **R** as $r_{i,j}$ and elements of a vector **r** as r_i. Derivatives with respect to time **t** are indicated using Leibniz's notation (e.g. $d\mathbf{R}/d\mathbf{t}$, $\partial\mathbf{R}/\partial\mathbf{t}$) or Newton's notation (e.g. $\dot{\mathbf{R}}$). Scalar variances, e.g. σ_r^2 for vector **r**, are self-explanatory.

Multidimensional arrays and tensors

Symbol	Dimension	Units	Description
$\underline{\mathbf{A}}$	(*ns* x 1 x *nw*)	$L \cdot mol^{-1}$	Tensor of pure component spectra
$\underline{\mathbf{C}}$	(nt_r x 1 x *ns*)	$mol \cdot L^{-1}$	Tensor of concentration profiles
$\underline{\mathbf{E}}$	(1 x nt_e x *ns*)		Tensor of elution profiles
$\underline{\mathbf{Y}}$	(nt_r x nt_e x *nw*)		Tensor of absorbance data

Matrices

Symbol	Dimension	Units	Description
A	(*ns* x *nw*)	$L \cdot mol^{-1}$	Pure component spectra of all species ($\ell\boldsymbol{\varepsilon}$)
$\mathbf{A_c}$	(*ns–nu* x *nw*)	$L \cdot mol^{-1}$	Fitted component spectra of the coloured species
$\mathbf{A_{uk}}$	(*ns–nks* x *nw*)	$L \cdot mol^{-1}$	Fitted component spectra of the species for which no known spectra have been provided
$\mathbf{B_A}$	(*nt* x *ns*)	$L^{-1} \cdot mol$	Inverse model calibration for **A**
$\mathbf{B_C}$	(*ns* x *nw*)	$mol \cdot L^{-1}$	Inverse model calibration for **C**
C	(*nt* x *ns*)	$mol \cdot L^{-1}$	Concentration profiles of all species
$\mathbf{C_c}$	(*nt* x *ns–nu*)	$mol \cdot L^{-1}$	Concentration profiles of coloured species
$\mathbf{C_{dos}}$	(*nt* x *ns*)	$mol \cdot L^{-1}$	Matrix of time resolved dosing concentrations
$\mathbf{C_{in}}$	(*nf* x *ns*)	$mol \cdot L^{-1}$	Matrix of time invariant dosing concentrations
$\mathbf{C_{uk}}$	(*nt* x *ns–nks*)	$mol \cdot L^{-1}$	Concentration profiles of the species for which no pure component spectra have been provided
$\mathbf{C_0}$	(*nt* x *ns*)	$mol \cdot L^{-1}$	Matrix of initial concentrations
$\mathbf{C_0^{ne}}$	(*ne* x *ns*)	$mol \cdot L^{-1}$	Initial concentrations of the *ne* additional experiments

Matrices (continuing)

Symbol	Dimension	Units	Description
D	(*nt* x *nt*)		Diagonal matrix of dilution
E	(*nr* x *ns*)		Reactant coefficients
H	(*nr* x *nr*)		Hessian
J	(*nt*·*nw* x *nr*)		Jacobian
N	(*nr* x *ns*)		Stoichiometric coefficients (**P** – **E**)
P	(*nr* x *ns*)		Product coefficients
R	(*nt* x *nw*)		Residuals
X	(*nt* x *nr*)	$mol \cdot L^{-1}$	Extents of reaction expressed in concentration units
Y	(*nt* x *nw*)		Time resolved multivariate spectroscopic data
Δ	(*ns*–*nu*–*nks* x *ns*)		Coefficient matrix defining the linear combinations of true (resolved) pure component spectra [1)]
ε	(*n*s x *nw*)	$L \cdot mol^{-1} \cdot cm^{-1}$	Molar absorptivities
ξ	(*nt* x *nr*)	mol	Extents of reaction expressed in mole units
Φ	(*nt* x *nr*)		Matrix of rate laws
Ω	(*ns*+*nf*+*ne*+1 x *ns*–*nu*–*nks*)		Time invariant augmented matrix pseudo-equivalent to **C**
$\mathbf{\Omega_{ind}}$	(nr_i+*nf*+*ne*+1 x *ns*–*nu*–*nks*)		Time invariant matrix with linearly independent rows

[1)] Symbol **Δ** is also used in Appendix 4.6.5 (calorimetry) with dimensions (*nr*–nu_r–*nke* x *nr*–*nke*) to indicate the coefficients of linear combinations of true enthalpies.

Vectors

Symbol	Dimension	Units	Description
$\mathbf{c_0}$	(1 x *ns*)	$mol \cdot L^{-1}$	Vector of initial concentrations
$\mathbf{c_0^{exp_i}}$	(1 x *ns*)	$mol \cdot L^{-1}$	Vector of initial concentrations of the *i*-th additional experiment (*i* = 2, 3, … , *ne*+1)
f	(*nt* x 1)	$mL \cdot min^{-1}$	Dosing (flow) rates ($\dot{\mathbf{v}}$)
$\mathbf{f_s}$	(1 x *nf*)	$mL \cdot min^{-1}$	Dosing (flow) rates of the different dosing steps
k	(1 x *nr*)	[1)]	Rate constants
$\mathbf{k_{opt}}$	(1 x *nr*)	[1)]	Optimised rate constants
p	(1 x *np*)		Parameters to optimise
$\dot{\mathbf{q}}_\mathbf{r}$	(*nt* x 1)	W	Reaction heat power
$\dot{\mathbf{q}}_\mathbf{flow}$	(*nt* x 1)	W	Heat power through the reactor wall
$\dot{\mathbf{q}}_\mathbf{comp}$	(*nt* x 1)	W	Heat power of the compensation heater
$\dot{\mathbf{q}}_\mathbf{lid}$	(*nt* x 1)	W	Heat power lost through the lid
$\dot{\mathbf{q}}_\mathbf{dos}$	(*nt* x 1)	W	Heat power of dosing
$\dot{\mathbf{q}}_\mathbf{cooling}$	(*nt* x 1)	W	Heat power flowing out of the reactor
$\dot{\mathbf{q}}_\mathbf{loss}$	(*nt* x 1)	W	Heat power lost to the environment
$\dot{\mathbf{q}}_\mathbf{pelt}$	(*nt* x 1)	W	Heat power pumped by peltier elements
$\dot{\mathbf{q}}_\mathbf{ohm}$	(*nt* x 1)	W	Heat power dissipated by peltier elements
$\dot{\mathbf{q}}_\mathbf{pelt}^\mathbf{loss}$	(*nt* x 1)	W	Heat power lost by conduction through peltier elements
r	(*nt*·*nw* x 1)		Vectorised residuals
T	(*nt* x 1)	K	Temperature
t	(*nt* x 1)	s	Reaction times
v	(*nt* x 1)	L	Time resolved reaction volumes
$\mathbf{x_0}$	(1 x *nr*)	$mol \cdot L^{-1}$	Initial extents of reaction
$\mathbf{\Delta_r H}$	(*nr* x 1)	$J \cdot mol^{-1}$	Reaction enthalpies
$\mathbf{\Delta_r H_c}$	(*nr*–nu_r–*nke* x 1)	$J \cdot mol^{-1}$	Reaction enthalpies of the contributing reactions
Δk	(1 x *nr*)	[1)]	Shift vector applied on **k**

[1)] Units depend on the order of the rate law ($[mol \cdot L^{-1} \cdot s^{-1}]$ for 0^{th}, $[s^{-1}]$ for 1^{st}, $[L \cdot mol^{-1} \cdot s^{-1}]$ for 2^{nd} and $[L^2 \cdot mol^{-2} \cdot s^{-1}]$ for 3^{rd} order rate constants)

Vectors (continuing)

Symbol	Dimension	Units	Description
$\boldsymbol{\theta}$			Spectral parameters
$\boldsymbol{\sigma}_{\mathbf{k}}^2$	(1 x *nr*)	[2)]	Variances in the rate constants
$\boldsymbol{\sigma}_{\mathbf{c_0}}^2$	(1 x *ns*)	$mol^2 \cdot L^{-2}$	Variances in the initial concentrations
$\boldsymbol{\sigma}_{\mathbf{f_s}}^2$	(1 x *nf*)	$L^2 \cdot s^{-2}$	Variances in the dosing (flow) rates $\mathbf{f_s}$
$\boldsymbol{\sigma}_{\mathbf{k,r}}^2$	(1 x *nr*)	[2)]	Variances of **k** due to the residuals
$\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$	(1 x *nr*)	[2)]	Variances of **k** due to the initial concentrations
$\boldsymbol{\sigma}_{\mathbf{k},\mathbf{f_s}}^2$	(1 x *nr*)	[2)]	Variances of **k** due to the dosing rates $\mathbf{f_s}$
$\boldsymbol{\sigma}_{\mathbf{k,T}}^2$	(1 x *nr*)	[2)]	Variances of **k** due to the fluctuations in temperature

[2)] Units depend on the order of the rate law ([$mol^2 \cdot L^{-2} \cdot s^{-2}$] for 0^{th}, [s^{-2}] for 1^{st}, [$L^2 \cdot mol^{-2} \cdot s^{-2}$] for 2^{nd} and [$L^4 \cdot mol^{-4} \cdot s^{-2}$] for 3^{rd} order rate constants)

Special matrices and vectors

Symbol	Description
$\mathbf{1}$	Matrix or vector of ones of appropriate dimensions
$\mathbf{I}$	Identity matrix of appropriate dimensions
$\mathbf{0}$	Matrix or vector of zeros of appropriate dimensions
$\mathbf{S}$	Matrix of all singular values (from SVD)
$\overline{\mathbf{S}}$	Matrix of significant singular values (from SVD)
$\mathbf{T}$, $\mathbf{T_{lin}}$, $\mathbf{T'_{lin}}$	Linear transformation matrices of appropriate dimensions
$\mathbf{U}$	Matrix of all orthonormal column eigenvectors (from SVD)
$\overline{\mathbf{U}}$	Matrix of significant orthonormal column eigenvectors (from SVD)
$\mathbf{V}$	Matrix of all orthonormal row eigenvectors (from SVD)
$\overline{\mathbf{V}}$	Matrix of significant orthonormal row eigenvectors (from SVD)
$\boldsymbol{\Theta}$	Diagonal or triangular matrix of appropriate dimensions
$\boldsymbol{\Lambda}$	Matrix of all eigenvalues ($\mathbf{S}^2$)

Operators

Symbol	Description
dim	Dimension of a matrix
im	Image of a matrix
ker	Kernel of a matrix
rank	Rank of a matrix
diag	Extracts a vector of diagonal elements from a matrix
DIAG	Generates a diagonal matrix from a vector
$\left\| \cdot \right\|^2$	Least squares fitting
$\min_{\mathbf{k}} \mathbf{R}$, $\min_{\boldsymbol{\theta}} \mathbf{R}$, $\min_{\mathbf{C}} \mathbf{R}$, $\min_{\mathbf{A}} \mathbf{R}$	Minimise **R** by changing **k**, **θ**, **C** or **A**
$\bullet\mathbf{E}^T$	Element-wise raise to the power of $\mathbf{E}^T$
$\otimes$	Tensor product
/ or \	Matlab linear regression (right or left division)
~	Pseudo-equivalence between two matrices
+ (superscript)	Moore-Penrose pseudo-inverse

Scalars

Symbol	Units	Description
$c_{0,A}$	mol·L^{-1}	Initial concentration of species *A*
$c_{in,A}$	mol·L^{-1}	Dosing concentration of species *A*
$c_{t,A}$	mol·L^{-1}	Concentrations of species *A* at time *t*
c_{tot}	mol·L^{-1}	Sum of initial concentrations
$c_{0,A}^{flask}$	mol·L^{-1}	Initial concentrations of species *A* in a flask
$d_{i,i}$		*i*-th diagonal element of the inverted Hessian $\mathbf{H}^{-1}$
E	J	Energy in Joules
E_{eV}	eV	Energy in electron-volts (1 eV = 1.60·10^{-19} J)
I	W·m^{-2}	Light intensity
k'		k_3/k_2 (Section 4.4.3)
ℓ	cm	Path length of the measurement cell
m_A	g	Weighed mass of *A*
M_A	g·mol^{-1}	Molecular mass of *A*
nc		Rank of matrix **C**
ne		Number of additional kinetic experiments
nf		Number of different dosing steps
nke		Number of a priori known enthalpies
nks		Number of a priori known pure component spectra
nr		Number of reactions
nr_i		Number of independent reactions
ns		Total number of species
nt, nt_r		Number of reaction times
nt_e		Number of elution times
nu		Number of uncoloured species
nu_r		Number of non-contributing reactions
nw		Number of wavelengths/wavenumbers
nw_{ex}, nw_{em}		Number of excitation/emission wavelengths
ssq		Sum or squared residuals
t, t_r	s	Reaction time
t_X	s	Reaction time required to reach a conversion *X*
v_0	L	Initial volume (v_1)
v_t	L	Volume at time *t*
v_A^{flask}	L	Volume of the flask containing species *A*
v_A^{pip1}	L	Volume of species *A* delivered by a 10 mL pipette
v_A^{pip2}	L	Volume of species *A* delivered by a 5 mL pipette
v_{tot}	L	$v_A^{pip1} + v_B^{pip1} + v_A^{pip2} + v_B^{pip2}$
x	m	Distance
X		Conversion
α, β, γ		Different ratios of initial concentrations
α', β', γ'		Different ratios of initial concentrations
δk_i, $\delta c_{0,n}$ and $\delta f_{s,m}$		Finite difference applied to k_i, $c_{0,n}$ and $f_{s,m}$
η_c		Refracting index of the ATR crystal
η_s		Refracting index of the sample
θ		Incident angle of an electromagnetic radiation
λ	m	Wavelength expressed in meters
$\lambda_{Å}$	Å	Wavelength expressed in Angströms (1 Å = 10^{-10} m)
λ_{nm}	nm	Wavelength expressed in nanometers (1 nm = 10^{-9} m)
$\lambda_{\mu m}$	μm	Wavelength expressed in micrometers (1 μm = 10^{-6} m)
μ		Any positive number different from 0 or 1
ν	Hz or s^{-1}	Frequency expressed in Hz
ν_{GHz}	GHz	Frequency expressed in GHz (1 GHz = 10^9 Hz)
$\bar{\nu}$	cm^{-1}	Frequency expressed in wavenumber

Chemical species

Symbol	Description
A, *Aa*, *AC*, *B*, *BP*, *C*, *D*, *E*, *P*, *S*	Chemical species
'*A*', '*Aa*', '*AC*', '*B*', '*BP*', '*C*', '*D*', '*E*', '*P*', '*S*'	Coloured species

Physical constants

Symbol	Value	Units	Description
c	$3 \cdot 10^{8}$	$m \cdot s^{-1}$	Speed of light
e	$1.60 \cdot 10^{-19}$	C	Elementary charge
h	$6.63 \cdot 10^{-34}$	J·s	Planck's constant

Short notation

$\boldsymbol{\Omega}^{\text{a) / b)}}_{\text{c) / d)}}$	a)	Uncoloured species in strategy (1), ∅: no uncoloured species
	b)	Pure component spectra in strategy (2), ∅: no pure component spectra provided
	c)	Dosed species in strategy (3), ∅: batch conditions
	d)	Varied initial concentrations in strategy (4), ∅: no additional experiment performed

ACRONYMS

Acronym	Description
ALS	Alternating Least Squares
ATR	Attenuated Total Reflectance
CFD	Computational Fluid Dynamics
CRC.v4	Combined Reaction Calorimeter Version 4
DFT	Density Functional Theory
EFA	Evolving Factor Analysis
FA	Factor Analysis
FSMW/EFA	Fixed-Size Moving Window/Evolving Factor Analysis
FT	Fourier Transform
GA	Genetic algorithm
IR	Infra-Red
ITTFA	Iterative Target Transformation Factor Analysis
MCR	Multivariate Curve Resolution
MCT	Mercury-Cadmium-Telluride (detector)
MW	Micro-Wave
NGL/M	Newton-Gauss Levenberg/Marquardt (algorithm)
NIPALS	Non-linear Iterative Partial Least Squares
NN	Neural Networks
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDE	Partial Differential Equation
PLS	Partial Least Squares
RFA	Resolving Factor Analysis
RREF	Reduced Row Echelon Form
SHE	Safety, Heath and Environment
SMCR	Self-Modelling Curve Resolution
SVD	Singular Value Decomposition
TFA	Target Factor Analysis
TT	Target Testing
TTFA	Target Transformation Factor Analysis
UV-vis	Ultra-Violet Visible
XR	X-Rays

CHAPTER 1

Introduction

1.1 Overview

In recent years, fine chemical and pharmaceutical industries have been forced to drastically reduce their costs to cope with globalisation and market liberalisation [1,2]. Current economic environment calls for reduced development times so that the duration of the competitive advantage provided by patents is extended [3]. Product quality and process monitoring become increasingly important, especially when patents have expired and new competitors enter the market [4-6]. Industry is also faced with increasing regulations related to safety, heath and environmental (SHE) aspects [7,8].

In parallel, analytical chemistry has considerably progressed in the last decade [9,10]. Multivariate analytical probes have been introduced, which are capable to follow in-situ the temporal evolution of chemical reactions and to identify some important intermediates [11-13]. This has allowed to elucidate more complex kinetic models [14,15] and the subsequent optimisation of the reaction conditions [16,17]. These instrumental and analytical developments have certainly been driven by the needs of industry for cheap and small-sized multivariate probes.

1.2 Motivation

Chemometric methods play an important role in the elaboration of kinetic models, as they allow extracting maximum information from the large amount of data produced by analytical instruments [18-21]. Kinetic hard-modelling is one of these methods, which consists in postulating a first-principles kinetic hard model for the underlying reaction and optimising the corresponding linear (e.g. absorptivity spectra) and non-linear (e.g. rate constants) parameters [22].

A gradient-based method is usually chosen for the task of optimisation and allows estimating the uncertainties in the optimised non-linear parameters [23]. The validity of the postulated model is assessed by statistical and chemical means, and the kinetic model is refined until a sufficient agreement with the data is obtained. Kinetic hard-modelling can be applied to various signals, e.g. spectroscopy [24], calorimetry [25], gas uptake and pressure [26].

Problem 1: Prediction of the uncertainties in the non-linear parameters

When a gradient-based method is used for optimisation, uncertainties in the optimised non-linear parameters are estimated using the variance/covariance matrix, which is generally based on the data only [27]. The drawback of this calculation method is that it systematically underestimates the uncertainties when compared to the ones resulting from multiple repetitions of the experiment under the same conditions. One important reason is that uncertainties in the experimental conditions, such as in the initial concentrations and in the dosing rate, have not yet been propagated into the errors of the non-linear parameters. In addition, the experimental conditions leading to minimum uncertainties in the optimised non-linear parameters cannot be predicted.

Problem 2: Analysis and validation of rank-deficient models

Linear dependencies in the concentration matrix of rank-deficient models represent a mathematical dilemma when spectroscopic data following Beer's law are directly fitted by kinetic hard-modelling, as the pseudo-inverse of the concentration matrix is not defined [28]. Different strategies exist to circumvent or break linear dependencies in the concentration matrix [27,29]. However, no method has been developed so far to elucidate the species involved in linear dependencies. This is so, even though experience and intuition already shows that rank deficiencies depend on the initial concentrations and derive from linear dependencies in the stoichiometry or in the rate law. Also, no framework has been proposed yet to systematically apply these different strategies and to predict the corresponding level of spectral resolution. The absence of a systematic method particularly causes difficulties for the spectral validation of rank deficient models.

1.3 Objectives

This doctoral thesis contributes to the development of chemometric methods for the kinetic hard-modelling of multivariate spectroscopic data. It has the following two objectives:

Objective 1: Solution to Problem 1

Development of a method for the propagation of uncertainties in the experimental conditions, e.g. uncertainties in the initial concentrations or in the dosing rate, into the errors of the non-linear parameters. This method leads to the prediction of experimental conditions minimising the uncertainties in the fitted non-linear parameters.

Objective 2: Solution to Problem 2

Elaboration of a method for the elucidation of the linear dependencies in the concentration profiles of rank deficient kinetic models and for the systematic determination of mathematical pre-treatment, a priori information, or experimental design, to be applied in order to optimise non-linear and linear parameters. This systematic method is based on a time invariant matrix covering rank deficiency due to the stoichiometry and to the rate law. It allows the spectral validation of the postulated kinetic model.

Both methods described in Objective 1 and Objective 2 are applied to simulated and experimental spectroscopic data.

Experimental data are obtained by following the reaction of benzophenone and phenylhydrazine under acetic acid catalysis leading to the corresponding hydrazone. The reaction is performed in the well-defined environment of a reaction calorimeter that is used to maintain highly isothermal conditions. The reaction path for the condensation of primary and secondary amines of hydrazine derivatives on carbonyl groups is presented in Section 1.5 (Appendix). Hydrazone derivatives are used in synthetic [30,31] and analytical [32,33] chemistry. They enter the manufacturing of various products, such as dyes [34], UV-absorbers, herbicides and insecticides [35]. More recently, hydrazones have found applications in biology [36] and in medicine, for example in the treatment of tuberculosis [37] and cancer [38].

1.4 Structure

This doctoral thesis is divided into six chapters. Note that Chapters 3, 4 and 5, and corresponding Appendices have been published in *Chemometrics and Intelligent Laboratory Systems* prior to the publication of this thesis.

Introduction (Chapter 1)

This first chapter describes the recent trends in chemical industry and in analytical chemistry. It also introduces the motivation and the objectives of the present thesis. Section 1.5 (Appendix) contains supplementary information related to the chemistry of hydrazones and their use.

Instrumental and chemometric tools in chemical kinetics (Chapter 2)

In the second chapter, experimental and chemometric tools used in chemical kinetics are introduced. Described experimental tools include mid-IR and UV-vis ATR-spectroscopy as well as calorimetry. Note that calorimetry is discussed on a theoretical basis only as the reaction followed in this thesis does not produce significant amounts of heat. Important chemometric methods relevant to chemical kinetics are reviewed using the distinction between hard- and soft-modelling. Section 2.3 (Appendix) contains additional information.

Uncertainties and error propagation in kinetic hard-modelling (Chapter 3)

Rigorous propagation of uncertainties in initial concentrations and in dosing rates into the errors in the fitted rate constants is discussed in this third chapter. The method is presented using simulated and experimental data. Conclusions are drawn regarding the optimal stoichiometric conditions to be used in order to determine second order rate constants. Additional information is given in Section 3.7 (Appendix).

Prediction of linear dependencies in the concentration profiles (Chapter 4)

The systematic prediction of linear dependencies in the concentration profiles is presented in the fourth chapter. Its implications on the kinetic hard-modelling of spectroscopic data are discussed using simulated data. Section 4.6 (Appendix) contains supplementary information related to this chapter, in particular to linear dependencies encountered in calorimetry (Section 4.6.5).

Spectral validation of rank deficient spectroscopic data (Chapter 5)

In the fifth chapter, both methodologies developed in Chapters 3 and 4 are discussed on an experimental case study. Data are collected under various experimental conditions and are used for the spectral assessment of a kinetic rank deficient mechanism. In Section 5.7 (Appendix), additional comments are provided.

Conclusion and outlook (Chapter 6)

This last chapter gives a general conclusion and some outlook for future developments and improvements. Note that specific conclusions to each method are also drawn at the end of Chapters 3, 4 and 5.

1.5 Appendix

1.5.1 Condensation of carbonyls into hydrazone derivatives

The reaction path for the condensation of carbonyl derivatives (**1**) with the primary amine of hydrazine derivatives (**2**) leading to their corresponding hydrazones (**5**) is shown in Figure 1.1 [30].

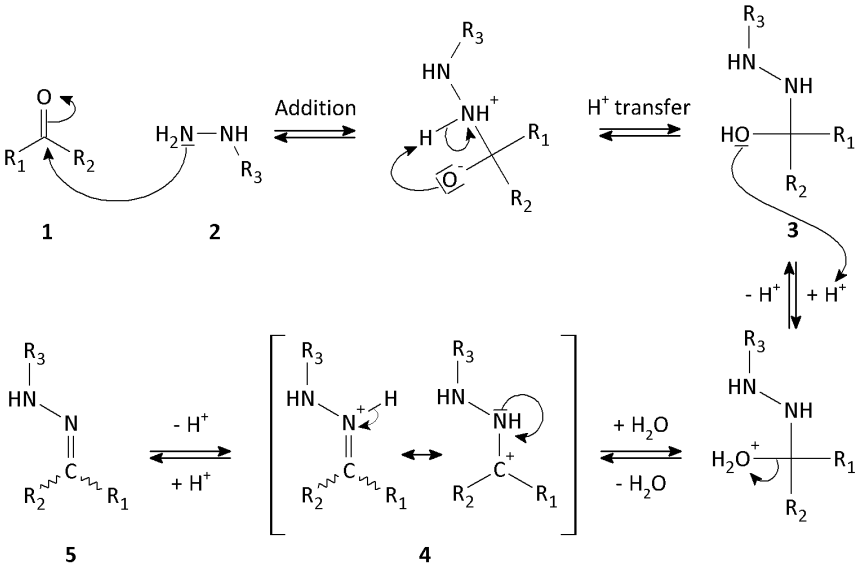

Figure 1.1: Reaction path for the condensation of carbonyl derivatives (**1**) with the terminal amine of hydrazine derivatives (**2**) forming an E- or Z- hydrazone (**5**). Stereoisomers (E and Z) of the hydrazone are not distinguished.

This condensation implies a nucleophilic addition by attack of the lone pair of the primary amine of hydrazine onto the carbonyl function, followed by a hydrogen transfer from the nitrogen to the oxygen, resulting in a hemiaminal derivative (**3**). The alcohol function of the hemiaminal is then protonated by a carboxylic acid (generally acetic acid, AcOH) added under catalytic concentration and forms, after spontaneous water elimination, an iminium derivative (**4**), which is stabilised by resonance. This iminium is finally deprotonated by the conjugated base of the carboxylic acid, resulting in two hydrazones that are stereoisomers (E- and Z- **5**, not distinguished in Figure 1.1), and allows the re-generation of the catalyst.

Note, when non-substituted hydrazines (**6**) are condensed with aromatic aldehydes (Ar = aryl), the second primary amine may subsequently condense, resulting in an azine (**7**) [31].

$$\underset{\mathbf{6}}{H_2N-NH_2} + 2\,ArCHO \xrightarrow{AcOH} \underset{\mathbf{7}}{ArCH=N-N=CHAr} + 2\,H_2O \qquad (1.1)$$

The reaction path described in Figure 1.1 is in agreement with the pseudo-second (third) order rate law postulated in Chapters 3 and 5, under batch (semi-batch) conditions, for the condensation of phenylhydrazine on benzophenone in THF and under acetic acid catalysis. Under the assumption that the bimolecular reaction of the ketone with hydrazine (first step of the reaction path) is the rate limiting step, the reaction path of Figure 1.1 can be reduced to a second order rate law under batch conditions. Under semi-batch conditions, the catalyst concentration has to be incorporated in the kinetic model, leading to a third order rate law, to account for the dilution and for the appearance of the dosed reactants initiating the reaction.

Another route to hydrazone derivatives is the aliphatic diazonium coupling, in which an acidic hydrogen of a methylene group is captured by an aqueous solution of sodium acetate. The resulting carbanion then reacts with a diazonium salt (ArN_2^+) to form the corresponding hydrazone [31].

In synthetic chemistry, hydrazones are used as a protection of the carbonyl function, as initial compound in the reduction of carbonyls into their corresponding alkanes (Wolff-Kishner reduction), and to obtain vinyl, indole, nitrile or amine compounds [30,31]. In analytical chemistry, it is employed to detect and isolate compounds containing a carbonyl group [32].

1.5.2 Condensation of carbonyls into enamine derivatives

The reaction path for the condensation of carbonyl derivatives (**8**) with the secondary amine of hydrazine derivatives (**9**) leading to their corresponding enamines (**11** and/or **12**) is shown in Figure 1.2 [30]. Condensation according to Figure 1.2 includes a nucleophilic addition and a proton transfer from the nitrogen to the oxygen, resulting in a hemiaminal derivative (**10**). The subsequent step, i.e. the water elimination, requires at least one aliphatic chain to be present on the α position of the carbonyl function.

When the carbonyl is substituted with one aliphatic chain, this results in one path and two products that are stereoisomers (E- and Z- **11** or E- and Z- **12**). When two aliphatic chains are present, two paths and four products are possible (E- and Z- **11** as well as E- and Z- **12**). Note that stereoisomers of the enamine are not distinguished in Figure 1.2.

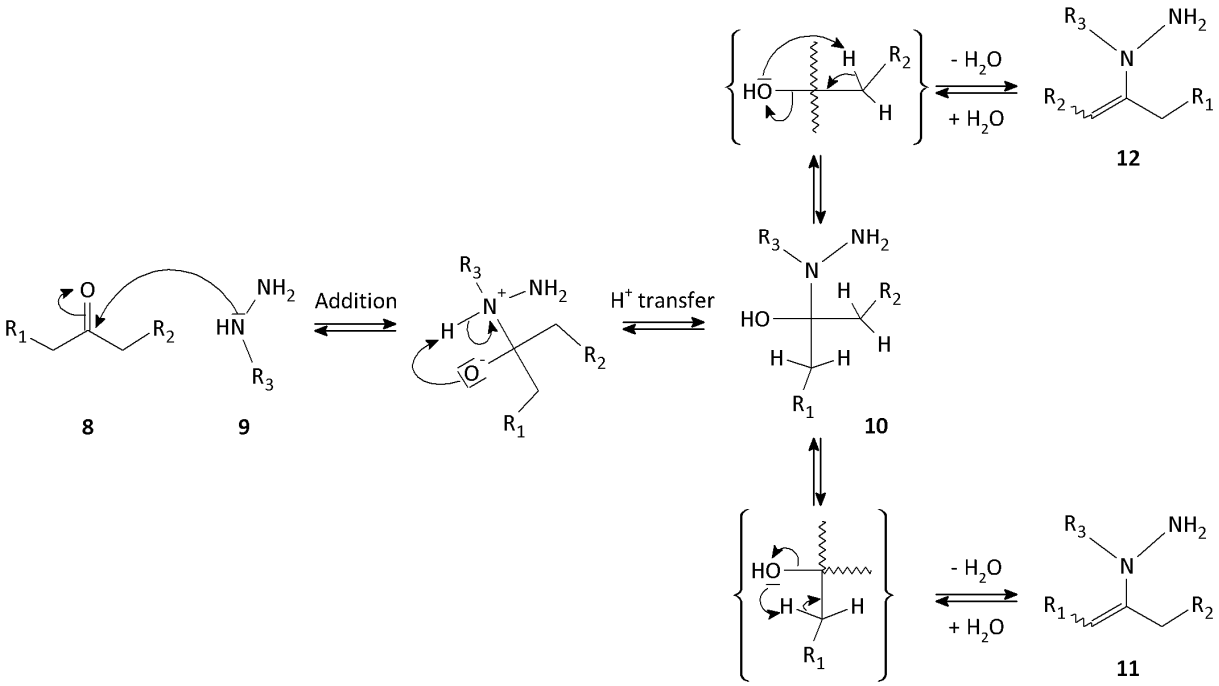

Figure 1.2: Reaction path for the condensation of carbonyl derivatives (**8**) with the secondary amine of hydrazine derivatives (**9**) forming enamines (**11** and/or **12**). Stereoisomers (E and Z) of the enamines are not distinguished.

As all hydrogens on the α position of the carbonyl compound used for the experiments of this thesis are aromatic (benzophenone), no enamine derivative can be formed. Thus the condensation of benzophenone with phenylhydrazine follows the reaction path described in Section 1.5.1 and results in the corresponding hydrazone.

1.6 References

[1] M.F. Edwards, Chemical Engineering Research & Design, 84 (2006), 255-260.
[2] R. Hussain, S. Wearne, Chemical Engineering Research & Design, 83 (2005), 372-378.
[3] E. Tirronen, T. Salmi, Chemical Engineering Journal, 91 (2003), 103-114.
[4] D.A. Obenndip, P.N. Sharratt, Chemical Engineering Research & Design, 83 (2005), 655-661.
[5] A.C. Quinn, P.J. Gemperline, B. Baker, M. Zhu, D.S. Walker, Chemom. Intell. Lab. Syst., 45 (1999), 199-214.
[6] E.N.M. van Sprang, H.J. Ramaker, H.F.M. Boelens, J.A. Westerhuis, D. Whiteman, D. Baines, I. Weaver, Analyst, 128 (2003), 98-102.
[7] E. Heinzle, K. Hungerbühler, Chimia, 51 (1997), 176-183.
[8] J. Workman, M. Koch, D. Veltkamp, Anal. Chem., 79 (2007), 4345-4363.
[9] T. Kourti, Crit. Rev. Anal. Chem., 36 (2006), 257-278.
[10] M. Spear, Chem. Proc., 70 (2007), 20-26.
[11] K.J. Laidler, J.H. Meiser, B.C. Sanctuary, Physical Chemistry, Fourth Edition, Houghton Mifflin, Boston, USA, 2003.
[12] H. Maskill (Ed.), Organic reactions and their mechanisms, Blackwell, Oxford, UK, 2006.
[13] R.M. Silverstein, G.C. Basler, T.C. Morill, Spectrometric identification of organic compounds, Fifth Edition, Wiley, New-York, USA, 1991.
[14] R.G. Wilkins, Kinetics and Mechanism of Reactions of Transition Metal Complexes, Second edition, VCH, Weinheim, Germany, 1991.
[15] J.H. Espenson, Chemical kinetics and reaction mechanisms, Second Edition, Mcgraw-Hill, New York, USA, 1995.
[16] T.F. Edgar, D.M. Himmelblau, L.S. Lasdon, Optimization of chemical processes, Second Edition, McGraw-Hill, Boston, USA, 2001.
[17] O. Levenspiel, Chemical Reaction Engineering, John Wiley & Sons, New York, USA, 1999.
[18] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Second Edition, Wiley, Weinheim, Germany, 2007.
[19] R.G. Brereton, Chemometrics - Data Analysis for the Laboratory and Chemical Plant, Wiley, England, 2003.
[20] P. Gemperline (Ed.), Practical Guide to Chemometrics, Taylor and Francis, Boca Raton, USA, 2006.
[21] E.R. Malinowski, Factor Analysis in Chemistry, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
[22] G. Puxty, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 81 (2006), 149-164.
[23] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C++ - The art of Scientific Computing, Second Edition, Cambridge University Press, New York, USA, 2005.
[24] G. Puxty, Y.-M. Neuhold, M. Jecklin, M. Ehly, P. Gemperline, A. Nordon, D. Littlejohn, J.K. Basford, M. De Cecco, K. Hungerbühler, Chem. Eng. Sci., 63 (2008), 4800-4809.
[25] A. Zogg, F. Stoessel, U. Fischer, K. Hungerbühler, Thermochim. Acta, 419 (2004), 1-17.
[26] G.D. Yadav, P.K. Goel, Clean Technologies and Environmental Policy, 4 (2003), 227-234.

[27] M. Maeder, Y.M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, NL, 2007.
[28] M. Maeder, A.D. Zuberbühler, Anal. Chem., 62 (1990), 2220-2224.
[29] Y.M. Neuhold, M. Maeder, J. Chemom., 16 (2002), 218-227.
[30] P.C. Vollhardt, N.E. Schore, Organic Chemistry: Structure and Function, Third Edition, W. H. Freeman and Company, New-York, USA, 1999.
[31] M.B. Smith, J. March, March's advanced organic chemistry: reactions, mechanisms, and structure, 5th Edition, John Wiley & Sons, Inc., New York, USA, 2001.
[32] R.B. Singh, P. Jain, R.P. Singh, Talanta, 29 (1982), 77-84.
[33] K. Andjelkovic, M. Sumar, I. Ivanovic-Burmazovic, J. Therm. Anal. Cal., 66 (2001), 759-778.
[34] A.M. Asiri, N.A. Fatani, Dyes and Pigments, 72 (2007), 217-222.
[35] M. Boger, D. Durr, L. Gsell, R.G. Hall, F. Karrer, O. Kristiansen, P. Maienfisch, A. Pascual, A. Rindlisbacher, Pest Management Science, 57 (2001), 191-202.
[36] S. Rollas, S.G. Kucukguzel, Molecules, 12 (2007), 1910-1939.
[37] A. Nayyar, R. Jain, Current Medicinal Chemistry, 12 (2005), 1873-1886.
[38] J.L. Buss, F.M. Torti, S.V. Torti, Current Medicinal Chemistry, 10 (2003), 1021-1034.

CHAPTER 2

Instrumental and chemometric tools in chemical kinetics

2.1 Instrumental tools

Calorimetry and spectroscopy are two instrumental techniques that are commonly used to follow the kinetics of chemical reactions and to determine their associated kinetic parameters, e.g. rate constants or activation energies. In addition to kinetic parameters, these two techniques also allow the determination of parameters inherently related to the measured signal, e.g. reaction enthalpies for calorimetry and absorptivity spectra for spectroscopy. Absorptivity spectra are particularly valuable parameters as they provide structural information on the reactive species suitable for data oriented chemical process development. Although not applied in this thesis, calorimetry is briefly discussed in this chapter as reactions of Chapter 3 and 5 were run in an isothermal reaction calorimeter.

2.1.1 Calorimetry

Calorimetry is a technique that is used to measure thermodynamic properties, e.g. reaction enthalpies, heat capacities and other thermal effects due to variations in physical properties, e.g. mixing or transition phase [1-5]. Calorimetry is also used to measure kinetic parameters, i.e. rate constants or activation energies [1,2]. Thermodynamic and kinetic information is particularly used for the purpose of safety assessment and reaction optimisation [3-7].

Calorimetric techniques can be classified by various ways, such as sample size (micro, small-scale, macro), sensitivity, temperature control (isothermal, adiabatic, isoperibolic), use of a reference sample (differential mode), measurement principle (heat-balance, heat-flow or heat compensation) and operational mode (batch, semi-batch, stirred). For a review on the classification of calorimetric techniques, we refer to [1].

All experiments performed in the present thesis (see Chapters 3 and 5) were carried out in the patented Combined Reaction Calorimeter (CRC.v4) [8,9], a 50 mL small-scale isothermal reaction calorimeter that combines the principles of power compensation and heat balance. In this calorimeter, power compensation is achieved by means of an electrical compensation heater made of hastelloy immersed into a hastelloy reactor stirred by a magnetically coupled stirrer. An ATR IR crystal is directly built into the bottom of the reactor and is connected to a FT-IR spectrometer via a system of conduits and mirrors (see Section 2.1.2.1). Six heat flow sensors, fixed around a hexagonal copper jacket, measure the heat flow from the reactor wall to the jacket by a heat balance principle, thus avoiding any calibration of heat transfer coefficients. The jacket temperature is kept constant by six peltier elements that pump heat in or out of the jacket. The pumped heat is compensated by a liquid circulating around the jacket, whose temperature is controlled by a thermostat. The reactor lid contains eight inserts, one of them being dedicated to an ATR UV-vis fibre optical probe (see Section 2.1.2.2). For a detailed description of the Combined Reaction Calorimeter, we refer to [9-14]. A schematic description of the CRC.v4 is shown in Figure 2.1.

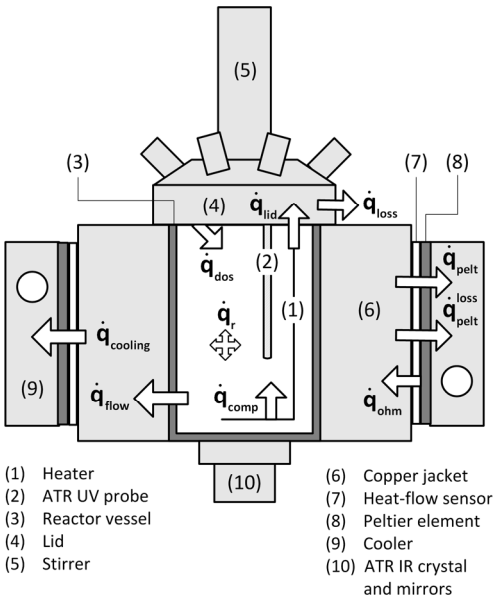

Figure 2.1. Schematic description of the CRC.v4 including the different contributions to the heat balance described in Equations 2.1 (heat balance on the reactor vessel), 2.2 (heat balance on the jacket) and 2.3 (combined heat balance).

An inner heat balance is applied to the reactor vessel and leads to Equation 2.1. An outer heat balance is also calculated for the jacket and gives an expression for the heat flow through the reactor wall as shown in Equation 2.2.

Heat balance on the reactor vessel: $$\dot{\mathbf{q}}_r = \dot{\mathbf{q}}_{flow} - \dot{\mathbf{q}}_{comp} + \dot{\mathbf{q}}_{lid} - \dot{\mathbf{q}}_{dos} \quad (2.1)$$

Heat balance on the jacket: $$\dot{\mathbf{q}}_{flow} = \dot{\mathbf{q}}_{cooling} - \dot{\mathbf{q}}_{lid} + \dot{\mathbf{q}}_{loss} \quad (2.2)$$

With $\dot{\mathbf{q}}_r$ being the reaction heat power, $\dot{\mathbf{q}}_{flow}$ the heat power dissipating through the reactor wall, $\dot{\mathbf{q}}_{comp}$ the heat power of the compensation heater, $\dot{\mathbf{q}}_{lid}$ the heat power transferred to the lid, $\dot{\mathbf{q}}_{dos}$ the heat power due to dosing, $\dot{\mathbf{q}}_{cooling}$ the overall cooling heat power flowing out of the reactor and $\dot{\mathbf{q}}_{loss}$ the calibrated loss of heat power. All described heat power (or heat flow) signals are univariate signals of dimensions (*nt* x 1) expressed in Watt, with *nt* denoting the number of reaction times.

Substituting $\dot{\mathbf{q}}_{flow}$ obtained from Equation 2.2 into Equation 2.1 results in an expression for the reaction heat power that does not require any calibration of the heat transfer coefficients through the reactor wall [9].

$$\dot{\mathbf{q}}_r = \dot{\mathbf{q}}_{cooling} - \dot{\mathbf{q}}_{comp} + \dot{\mathbf{q}}_{loss} - \dot{\mathbf{q}}_{dos} \quad \text{with} \quad \dot{\mathbf{q}}_{cooling} = \dot{\mathbf{q}}_{pelt} - \dot{\mathbf{q}}_{ohm} + \dot{\mathbf{q}}_{pelt}^{loss} \quad (2.3)$$

According to Equation 2.3, $\dot{\mathbf{q}}_{cooling}$ is calculated using $\dot{\mathbf{q}}_{pelt}$ the heat flow due to the peltier effect, $\dot{\mathbf{q}}_{ohm}$ the heat flow caused by the electrical resistance through the peltier elements and $\dot{\mathbf{q}}_{pelt}^{loss}$ the heat flow due to the conduction through the peltier elements. Alternatively, $\dot{\mathbf{q}}_{cooling}$ can directly be measured by heat flow sensors.

The reaction heat power $\dot{\mathbf{q}}_r$ is also related to the enthalpy of reaction according to Equation 2.4.

$$\dot{\mathbf{q}}_r = \dot{\boldsymbol{\xi}} \left(-\boldsymbol{\Delta}_r \mathbf{H}\right) \tag{2.4}$$

Where $\dot{\boldsymbol{\xi}}$ $(nt \text{ x } nr)$ denotes the derivative of the extent of reaction $\boldsymbol{\xi}$ (*nt* x *nr*) expressed in moles per unit of time, $\boldsymbol{\Delta}_r \mathbf{H}$ (*nr* x 1) the reaction enthalpies and *nr* the number of reaction steps. The derivative of the extent of reaction, $\dot{\boldsymbol{\xi}}$, is related to the rate law by Equation 2.5.

$$\dot{\boldsymbol{\xi}} = \mathrm{DIAG}(\mathbf{v})\, \boldsymbol{\Phi} = \mathrm{DIAG}(\mathbf{v})\, \dot{\mathbf{X}} + \mathrm{DIAG}(\dot{\mathbf{v}})\, \mathbf{X} \tag{2.5}$$

Where DIAG is an operator that generates a diagonal matrix from a vector argument, $\dot{\mathbf{v}}$ (*nt* x 1) denotes the vector of dosing rates **f** (*nt* x 1), **v** (*nt* x 1) the reaction volume, $\boldsymbol{\Phi}$ (*nt* x *nr*) the rate laws associated to each elementary reaction, and $\dot{\mathbf{X}}$ (*nt* x *nr*) the derivative of the extent of reaction, **X** (*nt* x *nr*), expressed in concentration per unit of time. Note that **X** (*nt* x *nr*) is sometimes referred as the degree of advancement [15]. The extent **X** in concentration units is related to the extent $\boldsymbol{\xi}$ in mole units by the reaction volume.

$$\boldsymbol{\xi} = \mathrm{DIAG}(\mathbf{v})\, \mathbf{X} \tag{2.6}$$

Due to its univariate nature, calorimetry is not sufficiently resolved for reliable kinetic model discrimination and process monitoring. Thus calorimetry is often combined with in-situ multivariate spectroscopy, which provides more detailed information to identify the reactive species and to elucidate the reaction mechanism [16-18].

2.1.2 Spectroscopy

Spectroscopic techniques are used in analytical chemistry for the identification and the quantification of chemical compounds [19]. In process development, multivariate (i.e. multi-wavelength) spectroscopy is a very powerful analytical technique to identify the reactive species, to elucidate the kinetic mechanism and to determine the associated kinetic parameters. Spectroscopy has also become the method of choice for process monitoring [20,21].

Spectroscopic techniques can be classified by various criteria such as the frequency range (UV-vis, mid-IR, microwave), the underlying spectroscopic transitions (Raman, fluorescence, phosphorescence, absorbance, emission), the principle of measurement (scanning spectroscopy, Fourier transform) or the spectroscopic probes (transmission or ATR probes).

In order to easily handle their characteristic ranges, spectroscopic techniques have traditionally different units for their multivariate axis, e.g. GHz for microwave spectroscopy, cm^{-1} or μm (microns) for vibrational spectroscopy, nm or eV (electron-volts) for UV-vis spectroscopy and Angström for X-rays [19]. Equation 2.7 shows the relationship between all these different units.

$$\underbrace{\nu = \nu_{GHz} 10^9 = (c\overline{\nu}) 10^2}_{\text{Frequency units}} = \underbrace{\left(\frac{c}{\lambda}\right) = \left(\frac{c}{\lambda_{Å}}\right) 10^{10} = \left(\frac{c}{\lambda_{nm}}\right) 10^9 = \left(\frac{c}{\lambda_{\mu m}}\right) 10^6}_{\text{Wavelength units}} = \underbrace{\left(\frac{E}{h}\right) = \left(\frac{E_{eV}}{h}\right) |e|}_{\text{Energy units}} \qquad (2.7)$$

Frequency units are indicated by ν, the frequency expressed in Hertz (s^{-1}), ν_{GHz}, the frequency in Gigahertz (microwave spectroscopy), and $\overline{\nu}$, the frequency expressed in wavenumber (vibrational spectroscopy). Wavelength units are denoted by λ, the wavelength expressed in meters, $\lambda_{Å}$, the wavelength in Angström (X-rays), λ_{nm}, the wavelength in nanometers (UV-vis spectroscopy) and $\lambda_{\mu m}$, the wavelength in microns (vibrational spectroscopy). Energy units are indicated by E, the energy in Joules, and E_{eV}, the energy in electron-volts. Physical constants are denoted by c, the speed of light ($3 \cdot 10^8$ $m \cdot s^{-1}$), h, Planck's constant ($6.63 \cdot 10^{-34}$ J·s), and $|e|$, the absolute value of the elementary charge of the electron, whose value is used to define the energy of one electron-volt ($1\ eV = |e|\ Joules = 1.60 \cdot 10^{-19}\ Joules$).

The different regions of the electromagnetic spectrum used in spectroscopy are shown in Figure 2.2.

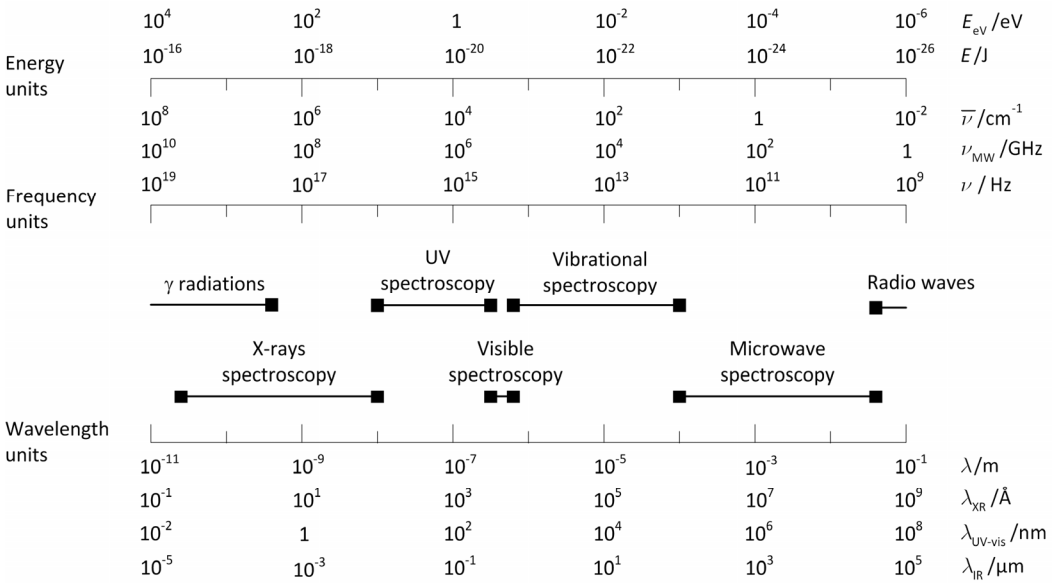

Figure 2.2. Electromagnetic spectrum and associated spectroscopic techniques described in energy, frequency and wavelength units.

Attenuated Total Reflectance (ATR) is a technology that can be used to measure in-situ the spectroscopic signal of chemical solutions, even in presence of solid particles. This is one of the advantages of ATR over transmission spectroscopy. In-situ spectroscopy (by ATR or by transmission) avoids the inherent drawbacks encountered in measurements in cuvettes, when offline samples have to be taken, e.g. changes of volume, of temperature, of pressure or perturbation due to the quenching method.

In ATR spectroscopy, the radiation passes through a transparent crystal with a high refracting index, allowing the radiation to be 'totally' reflected many times within the crystal (see Figure 2.3). Total reflectance occurs when the incident angle θ is higher than a certain critical angle that depends on the refracting index of the crystal (η_c) and of the sample (η_s).

Each time the light is reflected, a small part of the light is also absorbed by the solution and an evanescent wave is created at the contact between the crystal and the sample. This evanescent wave decays exponentially with the distance from the surface of the crystal. As the part of light that is absorbed is very small compared to the part that is reflected, ATR spectroscopy is preferred for the analysis of highly absorbing samples, i.e. compounds with high absorbing properties or highly concentrated solutions [22].

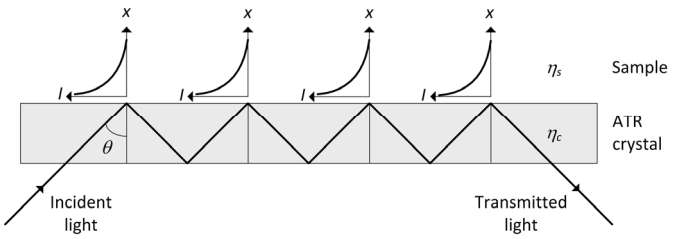

Figure 2.3. Incident light arriving at an angle θ and subjected to attenuated total reflectance (ATR), resulting in an attenuated transmitted light and an evanescent wave of intensity *I* at a distance *x*.

The depth of penetration of light into the sample due to the evanescent wave is defined as the distance from the crystal interface, for which the intensity of the evanescent wave is 37% of its initial intensity. This distance can be calculated using the refractive indices, the incident angle and the wavelength of the radiation, and is typically of the order of a few microns.

In spectroscopy, the measured signal is given as a transmittance (no units), i.e. the ratio between the transmitted light over the incident light (in $W \cdot m^{-2}$), or more frequently as an absorbance (no units), i.e. the negative logarithm of the transmittance. Beer's law is generally used for the quantification of spectroscopic measurements recorded by time and multivariate resolved spectroscopy.

In absence of molecular interaction between chemicals and under moderate absorbance, Beer's law states that the absorbances of a mixture of chemicals measured at *nw* wavelengths/wavenumbers and *nt* times, **Y** (*nt* x *nw*), is a linear combination of the path length of the measurement cell ℓ with the concentrations **C** (*nt* x *ns*) and the molar absorptivities **ε** (*ns* x *nw*) of all chemicals present in this mixture [19]. Matrix dimensions are indicated with *ns* the number of species.

As the path length is difficult to estimate in ATR spectroscopy, the product of molar absorptivities **ε** (*ns* x *nw*), which are normalised to unity concentration and unity path length ($L \cdot mol^{-1} \cdot cm^{-1}$), and the optical path ℓ, is usually replaced by pure component spectra, collected in matrix **A** (*ns* x *nw*), which define absorptivities normalised to unity concentration only ($L \cdot mol^{-1}$).

Thus for time and multivariate spectroscopic measurements, Beer's law can be written according to the following equation.

$$\mathbf{Y} = \mathbf{C}\ell\boldsymbol{\varepsilon} = \mathbf{CA} \qquad \text{with } \mathbf{A} = \ell\boldsymbol{\varepsilon} \tag{2.8}$$

2.1.2.1 mid-IR spectroscopy

Infrared spectroscopy covers the spectral domain between UV-vis and microwaves and can be divided into three regions: near-IR (14300 – 4000 cm^{-1}, close to the visible light), mid-IR (4000 – 400 cm^{-1}) and far-IR (400 – 10 cm^{-1}, close to the microwaves). The mid-IR range, corresponding to rotational-vibrational transitions of energy with a variation of dipole moment, is mainly used in chemistry [23]. Note that molecules, whose vibrations do not lead to a variation in their dipole moment, cannot be observed in mid-IR spectroscopy. They can, however, be observed in Raman spectroscopy, another technique of vibrational spectroscopy.

The mid-IR range can be further divided into three sub-regions: the high frequency sub-region where functional groups absorb (4000 – 1300 cm^{-1}, close to near-IR), the fingerprint sub-region (1300 – 900 cm^{-1}) and the low frequency sub-region (900 – 400 cm^{-1}, close to far-IR). As almost all compounds absorb in a different way in the fingerprint sub-region, mid-IR spectroscopy is particularly adapted for the identification of chemicals and for process monitoring [21]. This is also due to the sharpness of the peaks observed in mid-IR spectroscopy [23].

In the present work, mid-IR absorbances are recorded by a spectrometer (Mettler Toledo's ReactIR 4000, working range: 4000 – 650 cm^{-1}) using a Mercury-Cadmium-Telluride (MCT) detector that is cooled down to -195°C by liquid nitrogen, and working on the Fourier Transform (FT) principle. In FT mode, the incident radiation containing a spectrum of wavenumbers is divided into a fixed beam and a beam whose distance is adjusted by a mobile mirror. This mobile mirror is constantly adjusted in order to create an interferogram that can then be mathematically translated from the time domain into the frequency domain using the Fourier transformation. FT-IR spectrometers are generally faster and more sensitive than scanning spectrometers.

Molecular vibrations are sensitive to temperature, and so are the infrared pure component spectra. Therefore, in-situ infrared spectroscopic measurements are preferably recorded under isothermal conditions. For this purpose, the ReactIR 4000 spectrometer is connected via a K4 conduit (set of mirrors) to an ATR crystal built into the bottom of the vessel of an isothermal reaction calorimeter (see Section 2.1.1). In order to avoid adsorption of water on the mirrors, the K4 conduit is constantly flushed with dry nitrogen. The ATR crystal, a high pressure DiComp crystal made of ZnSe coated with diamond and surrounded by Hastelloy, is semi-transparent and absorbs in a defined region (between 2250 and 1950 cm^{-1}), not useful for analysis.

2.1.2.2 UV-vis spectroscopy

Ultraviolet spectroscopy covers the spectral domain between X-rays and visible spectroscopy and can be divided into two regions: the near-UV (200 – 380 nm, close to the visible light) and the far-UV (200 – 10 nm, close to the X-rays). Visible spectroscopy covers the visible part of light, i.e. between 380 (blue light) and 800 nm (red light). Near-UV and visible ranges are mostly used in chemistry and are commonly combined under the terminology UV-vis spectroscopy. Absorption in the UV-vis range essentially occurs in conjugated chemical systems, i.e. systems with alternating single and double bonds (e.g. butadiene, aromatic compounds), where valence electrons are promoted from low to high energetic molecular orbitals (electronic transitions) and result in high transition dipole moments [23].

Absorption bands in UV-vis are broad and relatively uncharacteristic. Thus the sole information obtained on the UV-vis range is not always sufficient to completely elucidate the structure of chemical compounds. That is why results from UV-vis spectroscopy are often interpreted in combination with other data obtained from other spectroscopic techniques, e.g. mid-IR spectroscopy.

In this thesis, UV-vis absorbances are recorded by a scanning spectrophotometer (Varian's Cary 50), with a xenon lamp as source and a photomultiplier as detector, measuring between 190 and 1100 nm. The Cary 50 spectrophotometer measures in-situ the reactor content using an ATR UV fibre-optical probe (Hellma) made of hastelloy inserted in the reactor vessel via a dedicated lid insert (see 2.1.1). Note that, in contrast to infrared spectroscopy, isothermal conditions are not considered as strict a requirement for UV-vis spectroscopy, since electronic transitions and UV-vis pure component spectra are less sensitive to temperature.

2.2 Chemometric tools for kinetic analysis

Various chemometric tools have been developed in order to obtain maximum information from the large amount of data produced by recent analytical instruments [24-27]; when dedicated to the resolution of multivariate signals, these chemometric tools are often referred to as Multivariate Curve Resolution (MCR) methods. Chemometric methods aiming for kinetic resolution can generally be classified by two distinct categories: kinetic hard-modelling and kinetic soft-modelling. These two categories are discussed in the following overview. Note that amongst the methods introduced below it was direct kinetic hard-modelling that was applied in this thesis.

2.2.1 Kinetic hard-modelling

Kinetic hard-modelling consists in modelling some part of the data using a chemical or physical first principles model and determining some kinetic parameters by optimisation. In bilinear spectroscopy and according to Beer's law (Equation 2.8), either concentration profiles (**C**, i.e. time profiles) or pure component spectra (**A**, i.e. wavelength profiles) can be modelled. In calorimetry, only the derivative of the reaction extents $\dot{\xi}$ is generally modelled (see Equation 2.4). Figure 2.4 (in Section 2.2.1.6) summarises the different approaches of kinetic hard-modelling discussed in Sections 2.2.1.1 – 2.2.1.5. For a description of data structure (including bilinearity) and measurement order, we refer to Sections 2.3.1 and 2.3.2 (Appendices).

2.2.1.1 Modelling C or A

In spectroscopy, **C** is modelled by a first principles model (the rate law) as a set of Ordinary Differential Equations (ODE) depending on a limited number of kinetic parameters, e.g. the rate constants **k** (1 x *nr*) of *nr* elementary reactions, to be determined by non-linear optimisation [28]. For a distinction between linear and non-linear parameters, see Section 2.3.3 (Appendix).

An alternative consists in modelling **A** as a set of Gaussian functions depending on spectral parameters θ [29]. The drawback of modelling **A** using this approach is the large number of spectral parameters to be defined and also that **A** does not directly depend on kinetic parameters. Thus, each time **A** is modelled, **C** has also to be modelled as a set of ODE depending on kinetic parameters to be fitted. In practice, modelling **A** (and thus **C**) requires the separation of the optimisation in two successive fitting procedures: first, the optimisation of the spectral parameters θ and the determination of concentration profiles **C**, and second, the optimisation of **k** by fitting these concentration profiles and by fixing the spectral parameters θ to their optimised values. This could be the reason why modelling **A** is not as frequent as modelling **C**.

Following Taavitsainen's convention [30,31], kinetic hard-modelling methods can be classified according to the fitted data (direct or indirect classification) and according to the way of determining **A** (when **C** is modelled) or **C** (when **A** is modelled) for spectroscopy, and of determining $\mathbf{\Delta_r H}$ ($\dot{\boldsymbol{\xi}}$ modelled) for calorimetry (implicit or explicit calibration).

2.2.1.2 Direct kinetic hard-modelling

In direct kinetic hard-modelling, the spectral measurement is expressed as a function of the modelled part of the data, $\mathbf{C}_{\mathbf{modelled}}$ or $\mathbf{A}_{\mathbf{modelled}}$ (see Equation 2.8), and the calorimetric measurement depends on $\dot{\boldsymbol{\xi}}_{\mathbf{modelled}}$ (see Equation 2.4). Note that the subscript 'modelled' indicates the part of the data that is modelled. The objective function used as driving force for the parameters optimisation is defined as the difference between the measured and the modelled spectral data.

For spectroscopy, two objective functions can be designed depending on which part of Beer's law is modelled.

$$\min_{\mathbf{k}} \left\| \mathbf{Y} - \mathbf{C}_{\mathbf{modelled}}\mathbf{A} \right\|^2 \qquad \text{when } \mathbf{C} \text{ is modelled [28]} \qquad (2.9)$$

$$\min_{\boldsymbol{\theta}} \left\| \mathbf{Y} - \mathbf{C}\mathbf{A}_{\mathbf{modelled}} \right\|^2 \text{ and } \min_{\mathbf{k}} \left\| \mathbf{C} - \mathbf{C}_{\mathbf{modelled}} \right\|^2 \qquad \text{when } \mathbf{A} \text{ is modelled [29]} \qquad (2.10)$$

For the sake of simplicity, the square double bar notation ($\| \cdot \|^2$) is introduced and indicates the sum of all squared elements of a matrix or vector argument. This results in an optimisation in the least-squares sense. Note that **C** in Equation 2.10 is obtained by performing $\mathbf{Y}\mathbf{A}^{+}_{\mathbf{modelled}}$ (see Section 2.2.1.4).

For calorimetry, only one objective function can be defined [2,16,18].

$$\min_{\mathbf{k}} \left\| \dot{\mathbf{q}}_{\mathbf{r}} - \dot{\boldsymbol{\xi}}_{\mathbf{modelled}} \left(-\mathbf{\Delta_r H} \right) \right\|^2 \qquad (2.11)$$

2.2.1.3 Indirect (inverse) kinetic hard-modelling

In indirect kinetic hard-modelling, **C** or **A** are expressed as a function of the spectral measurement **Y**. This hard-modelling technique uses an inverse model (see Equations 2.12 and 2.13) to relate the measurements to concentrations (or pure component spectra) rather than a direct model, such as Beer's law (Equation 2.8). Principal Component Regression (PCR) [32], Partial Least Squares (PLS) [25,33] and Neural Networks (NN) [34] are some inverse modelling techniques potentially applicable in indirect kinetic hard-modelling. Note that PCR and PLS methods require multivariate signals.

$$\mathbf{C} = \mathbf{Y}\mathbf{B}_{\mathbf{C}} \qquad \text{when } \mathbf{C} \text{ is modelled [32]} \qquad (2.12)$$

$$\mathbf{A} = \mathbf{B}_{\mathbf{A}}\mathbf{Y} \qquad \text{when } \mathbf{A} \text{ is modelled} \qquad (2.13)$$

Under ideal conditions, matrices $\mathbf{B_C}$ (*ns* x *nw*) and $\mathbf{B_A}$ (*nt* x *ns*) are equal to $\mathbf{A}^+$ and $\mathbf{C}^+$ respectively, where the superscript '+' denotes the Moore-Penrose pseudo-inverse operation. For details regarding the pseudo-inverse calculation, we refer to Section 2.3.4 (Appendix). In indirect kinetic hard-modelling, the objective function is defined as the difference between **C** (or **A**) and the inverse model of Equation 2.12 (or Equation 2.13).

$$\min_{\mathbf{k}} \left\| \mathbf{C}_{\mathbf{modelled}} - \mathbf{Y}\mathbf{B}_{\mathbf{C}} \right\|^2 \qquad \text{when } \mathbf{C} \text{ is modelled} \qquad (2.14)$$

$$\min_{\boldsymbol{\theta}} \left\| \mathbf{A}_{\mathbf{modelled}} - \mathbf{B}_{\mathbf{A}}\mathbf{Y} \right\|^2 \text{ and } \min_{\mathbf{k}} \left\| \mathbf{C} - \mathbf{C}_{\mathbf{modelled}} \right\|^2 \qquad \text{when } \mathbf{A} \text{ is modelled} \qquad (2.15)$$

Note that **C** in Equation 2.15 is obtained by performing $\mathbf{Y}\mathbf{A}^{+}_{\mathbf{modelled}}$ (see Section 2.2.1.4).

2.2.1.4 Implicit calibration

In direct kinetic hard-modelling by implicit calibration, Equations 2.8 (spectroscopy) and 2.4 (calorimetry) are used. When **C**, **A** or $\dot{\boldsymbol{\xi}}$ are modelled, the implicit calibration implies the use of the measurement itself (**Y** or $\dot{\mathbf{q}}_r$) at each iteration of the optimisation in order to determine respectively **A**, **C** or $\boldsymbol{\Delta}_r\mathbf{H}$ by linear regression. As determining the counter part of the data by this way does not require any additional external information, it is also deservedly referred as calibration free in literature [28,35].

For spectroscopy, this linear regression is performed using Equations 2.16 or 2.17. For more details regarding these equations, we refer to Section 2.3.4 (Appendix).

$$\mathbf{A} = \mathbf{C}_{\text{modelled}}^{+}\mathbf{Y} \qquad \text{when } \mathbf{C} \text{ is modelled [28,36,37]} \qquad (2.16)$$

$$\mathbf{C} = \mathbf{Y}\mathbf{A}_{\text{modelled}}^{+} \qquad \text{when } \mathbf{A} \text{ is modelled [29]} \qquad (2.17)$$

For calorimetry, this linear regression is performed using Equation 2.18 [16].

$$\mathbf{\Delta}_r\mathbf{H} = -\dot{\boldsymbol{\xi}}_{\text{modelled}}^{+}\,\dot{\mathbf{q}}_r \qquad (2.18)$$

In indirect kinetic hard-modelling by implicit calibration, Equations 2.12 and 2.13 are used when spectroscopic data are analysed. It is out of the scope of this work to describe the fairly complex steps required to obtain matrices $\mathbf{B_C}$ and $\mathbf{B_A}$ for Partial Least Squares [30,31,38,39]. For PCR, however, the procedure can be briefly described. Singular Value Decomposition (see 2.2.2.1) is first performed on matrix **Y** in order to obtain matrices **U**, **S** and **V**. Then, retaining only the significant part of these matrices (PCA, see 2.2.2.1), i.e. $\overline{\mathbf{U}}$, $\overline{\mathbf{S}}$ and $\overline{\mathbf{V}}$, matrix **Y** is reduced to $\overline{\mathbf{Y}} = \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}$ in Equation 2.12 or 2.13, and computing the pseudo-inverse of $\overline{\mathbf{Y}}$, i.e. $\left(\overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}\right)^{+}$, allows the determination of $\mathbf{B_C}$ or $\mathbf{B_A}$. Note that replacing **Y** by $\overline{\mathbf{Y}}$ avoids the calculation of the badly conditioned pseudo-inverse of **Y**.

$$\mathbf{B}_{\mathbf{C}} = \left(\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^{\mathrm{T}}\right)\mathbf{C}_{\text{modelled}} \qquad \text{when } \mathbf{C} \text{ is modelled [32]} \qquad (2.19)$$

$$\mathbf{B}_{\mathbf{A}} = \mathbf{A}_{\text{modelled}}\left(\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^{\mathrm{T}}\right) \qquad \text{when } \mathbf{A} \text{ is modelled} \qquad (2.20)$$

Equations 2.19 and 2.20 are obtained using the property of orthonormal vectors with respect to inversion: $\left(\overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}\right)^{+} = \overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^{\mathrm{T}}$. Note that left and right pseudo-inverses are equal in this case.

When **C** is modelled, the Target Testing (TT, see 2.2.2.2) method proposed by Furusjö et al [40], which projects **C** onto the $\overline{\mathbf{U}}$ eigenspace to reduce effects of baseline shifts during the fitting, can be regarded as an indirect kinetic hard-modelling by implicit calibration using PCR.

This is shown by reducing matrix **Y** to $\overline{\mathbf{Y}} = \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}$ in the objective function described in Equation 2.14 and replacing $\mathbf{B_C}$ by its expression derived in Equation 2.19.

$$\min_{\mathbf{k}} \left\| \mathbf{C}_{\mathbf{modelled}} - \left(\overline{\mathbf{U}}\overline{\mathbf{U}}^{\mathrm{T}}\right)\mathbf{C}_{\mathbf{modelled}} \right\|^2 \tag{2.21}$$

When **A** is modelled, the parallel approach is obtained by replacing matrix **Y** by $\overline{\mathbf{Y}} = \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}$ in the objective function described in Equation 2.15 and replacing $\mathbf{B_A}$ by its expression derived from Equation 2.20. The first objective function (for the optimisation of the spectral parameters θ) and the second objective function (for the optimisation of the kinetic parameters **k**) are given in Equation 2.22.

$$\min_{\boldsymbol{\theta}} \left\| \mathbf{A}_{\mathbf{modelled}} - \mathbf{A}_{\mathbf{modelled}}\left(\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{V}}\right) \right\|^2 \quad \text{and} \quad \min_{\mathbf{k}} \left\| \left(\overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}\right)\mathbf{A}^{+} - \mathbf{C}_{\mathbf{modelled}} \right\|^2 \tag{2.22}$$

Note that **A** in Equation 2.22 is the optimum $\mathbf{A}_{\mathbf{modelled}}$.

2.2.1.5 Explicit calibration

In direct kinetic hard-modelling by explicit calibration, the linear coefficients **A**, **C** or $\mathbf{\Delta_r H}$, are determined independently using a training set (also called calibration set). As an example, an explicit calibration in direct kinetic hard-modelling of spectroscopic data (modelling **C**) consists in directly measuring the *ns* pure component spectra collected in **A** [32,41]. In indirect kinetic hard-modelling by explicit calibration (modelling **C**), a training set is used to independently determine the linear coefficients in $\mathbf{B_C}$ [33]. Direct and indirect kinetic hard-modelling by explicit calibration are usually not the best option when **A** is modelled as determining **C** by calibration represents a tremendous task.

2.2.1.6 Non-linear optimisation algorithms

Gradient-based algorithms are very popular for non-linear optimisation of kinetic hard-modelling problems, as they converge quickly towards the minimum and allow estimating uncertainties in the fitted parameters. An example is the Newton-Gauss Levenberg/Marquardt (NGL/M) algorithm, which converges quadratically with the number of iterations and constantly adapts the optimisation step according to convergence (Newton-Gauss algorithm) or local divergence (Levenberg-Marquardt modification) [32,42]. For more details about this gradient-based algorithm, we refer to Chapter 3.

Non-gradient based algorithms directly use function evaluations (not derivatives) to guide the optimisation towards the minimum. Compared to gradient-based algorithms, gradient-free methods converge slowly and require many function evaluations at each iteration. In addition, they do not allow estimating uncertainties in the fitted parameters.

An example of a non-gradient based method is the simplex algorithm, which approaches the minimum by successively reflecting multi-dimensional polygons in the opposite direction given by the maximum corner [42]. Another example is the genetic algorithm (GA), which finds a population of optimised parameters by applying concepts borrowed from genetics (e.g. selection, mutations and crossover) to successive generations of parameter populations [17,43].

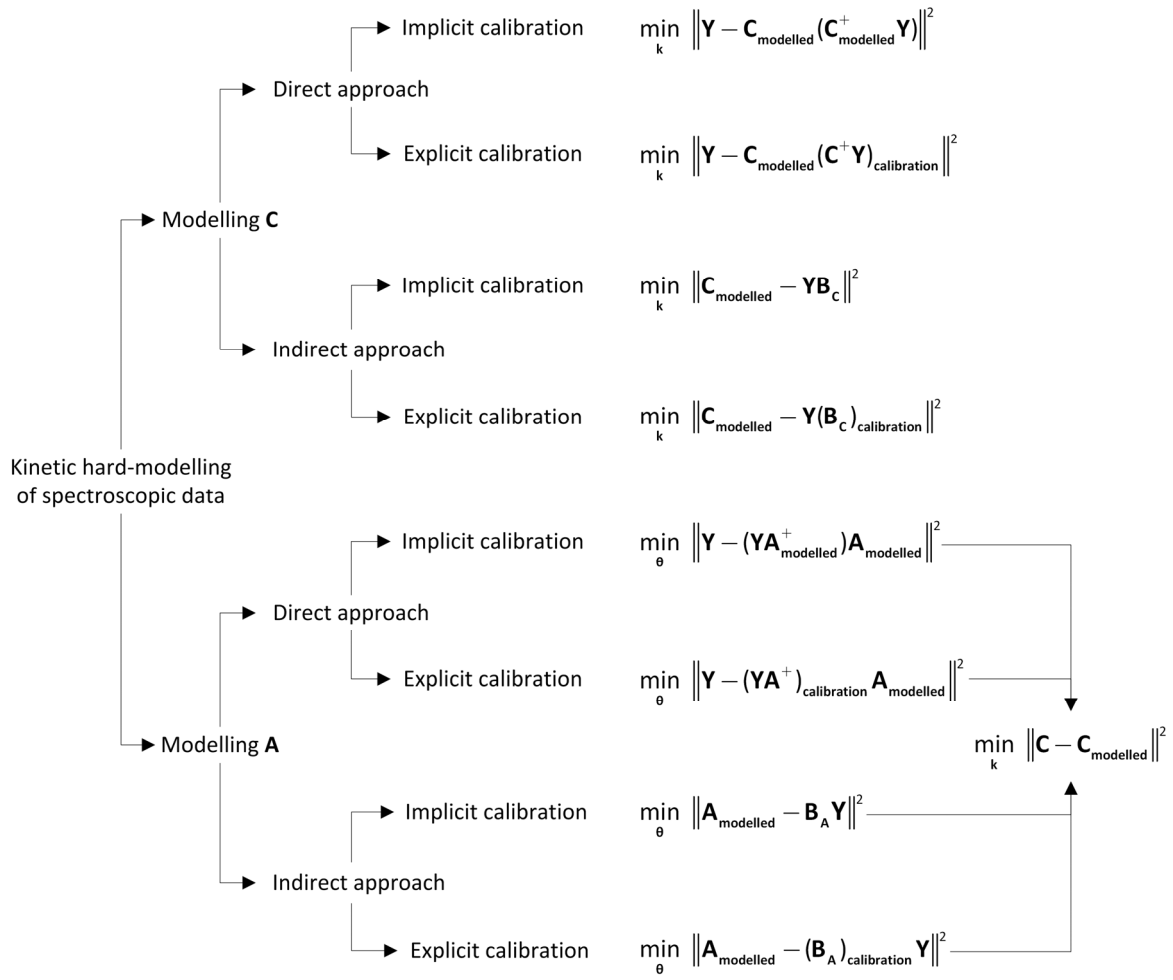

Figure 2.4. Decision tree for the selection of an appropriate method for the kinetic hard-modelling of spectroscopic data. Note that for the modelling of **A**, a subsequent hard-modelling of the concentrations **C** (obtained from $\mathbf{YA}^{+}$, where **A** is the optimum $\mathbf{A}_{modelled}$) is required.

2.2.2 Kinetic soft-modelling

Soft-modelling, also referred as Self-Modelling Curve Resolution (SMCR) [44], covers different model-free methods applicable to multivariate signals. Examples of soft-modelling methods are Factor Analysis (FA), Target Factor Analysis (TFA), Evolving Factor Analysis (EFA) and Alternating Least Squares (ALS). For a review on two-way and three-way soft-modelling techniques, see [45].

All kinetic soft-modelling methods applied to spectroscopy are devoted to the calculation of Beer's law matrices, **C** and **A**. As no kinetic hard model is postulated, no rate constants can be obtained by this means. In addition, all these methods suffer from the same inherent problem, namely the linear correlation between matrices **C** and **A** in Beer's law (Equation 2.8) that leads to a non-unique solution for the decomposition of **Y** into a product of two matrices. This problem is called Rotational Ambiguity [46]. For very simple reactions, applying soft-constraints (e.g. non-negativity or unimodality) can break this linear correlation, allowing the determination of **C** and **A** without any ambiguity, except for scaling factors. For more complex reactions, no general remedy exists to this rotational ambiguity [32].

2.2.2.1 Factor analysis

Factor Analysis (FA) was originally developed for analysing sources of variation (factors) in psychological and sociological data [27]. In Chemometrics, it was introduced in order to decompose a set of bilinear measurements into a product of matrices with particular properties and containing abstract information [32].

These last years, Singular Value Decomposition (SVD), a particular method (and algorithm) for Factor Analysis, has become the method of choice to extract information from multivariate spectral measurements. It relies on the decomposition of a data matrix **Y** (*nt* x *nw*) into a column-orthonormal matrix **U** (*nt* x *nw*), a diagonal matrix **S** (*nw* x *nw*) comprised of positive or zero elements, and a row-orthonormal matrix **V** (*nw* x *nw*) [27]. Dimensions for the decomposition are given here assuming $nw < nt$. For $nt < nw$, dimensions are reduced to (*nt* x *nt*) for **U**, (*nt* x *nt*) for **S** and (*nt* x *nw*) for **V** [42].

$$\mathbf{Y} = \mathbf{USV} \tag{2.23}$$

Another algorithm, called Non-linear Iterative Partial Least Squares (NIPALS), consists in writing matrix **Y** as the product of a score matrix and of a matrix of loadings [25,45]. The result of this decomposition is commonly displayed in so-called score plots. Comparison with Equation 2.23 shows that the score matrix is linearly dependent on **U** (or **US**) and the loading matrix is linearly dependent on **SV** (or **V**).

Columns of **U** and rows of **V** contain the eigenvectors of the square matrices $\mathbf{YY}^T$ and $\mathbf{Y}^T\mathbf{Y}$ respectively. Note the following properties for these matrices: $\mathbf{U}^+ = \mathbf{U}^T$, $\mathbf{V}^+ = \mathbf{V}^T$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{VV}^T = \mathbf{I}$, where **I** is an identity matrix of appropriate dimensions. Matrix **S** of dimensions (*nw* x *nw*) or (*nt* x *nt*) contain singular values which correspond to the square roots of the eigenvalues $\mathbf{\Lambda}$ of dimensions (*nw* x *nw*) or (*nt* x *nt*), i.e. $\mathbf{\Lambda} = \mathbf{S}^2$.

Singular values in **S** are usually ranked in descending order of magnitude along the diagonal, the smallest values related to so-called insignificant components having a marginal effect on the reconstitution of **Y** using Equation 2.23 and mainly accounting for the noise in **Y**. A significance level can be determined by statistical means (e.g. F-test) or by visual inspection of the singular values. Also, possible patterns in the abstract vectors **U** and **V** can be used to determine the number of significant components [27].

Removing the insignificant singular values from **S** and insignificant components from **U** (corresponding columns) and **V** (corresponding rows) allows reducing **Y** to its significant information. This procedure is called Principal Component Analysis (PCA) and results in matrices $\overline{\mathbf{U}}$ $(nt \text{ x } nc)$, $\overline{\mathbf{S}}$ $(nc \text{ x } nc)$ and $\overline{\mathbf{V}}$ $(nc \text{ x } nw)$, *nc* denoting the number of significant retained singular values (or eigenvectors) [25,27,32,45] (see Equation 2.24). This number is also the number of reactive species that can be distinguished in spectroscopy, i.e. the linearly independent rows or columns in matrix **Y**. Importantly, note that removing insignificant factors in **U** and **V** has the following consequence: $\overline{\mathbf{U}}\overline{\mathbf{U}}^T \neq \mathbf{I}$ and $\overline{\mathbf{V}}^T\overline{\mathbf{V}} \neq \mathbf{I}$.

As the decomposition is truncated after the first *nc* significant abstract factors, PCA results in a reduced **Y**, denoted $\overline{\mathbf{Y}}$ $(nt \text{ x } nw)$, ideally only comprising time and multivariate fluctuations of **Y** that are significantly different from noise.

$$\overline{\mathbf{Y}} = \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}} \tag{2.24}$$

2.2.2.2 Target factor analysis

Abstract eigenvectors in $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ obtained from PCA are linearly correlated to **C** and **A** respectively, matrix $\overline{\mathbf{S}}$ comprising weighting factors that can be applied either on $\overline{\mathbf{U}}$ or on $\overline{\mathbf{V}}$. As a convention in this work, $\overline{\mathbf{S}}$ will be attributed to $\overline{\mathbf{V}}$. The linear function relating abstract matrices ($\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$) to real matrices (**C** and **A**) can be derived by introducing a transformation matrix **T** in the decomposition of PCA (Equation 2.24) and by comparing this decomposition with Beer's law (Equation 2.8).

$$\overline{\mathbf{Y}} = \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}} = \overline{\mathbf{U}}\left(\mathbf{T}\mathbf{T}^{-1}\right)\overline{\mathbf{S}}\,\overline{\mathbf{V}} = \left(\overline{\mathbf{U}}\mathbf{T}\right)\left(\mathbf{T}^{-1}\overline{\mathbf{S}}\,\overline{\mathbf{V}}\right) = \mathbf{C}\mathbf{A} \tag{2.25}$$

Comparing the terms of Equation 2.25 allows writing the following relationship for **C** and **A**.

$$\mathbf{C} = \overline{\mathbf{U}}\mathbf{T} \tag{2.26}$$

$$\mathbf{A} = \mathbf{T}^{-1}\overline{\mathbf{S}}\,\overline{\mathbf{V}} \tag{2.27}$$

Equations 2.26 and 2.27 are the expressions of Target Factor Analysis (TFA) [27,32,45,47]. Different methods exist in order to calculate the transformation matrix, the most evolved method being Resolving Factor Analysis (RFA) where the elements of **T** are determined by non-linear optimisation [47].

A method that avoids calculating the transformation matrix **T** is Target Transformation Factor Analysis (TTFA). It consists in testing if matrices $\mathbf{C}_{\mathbf{tested}}$ or $\mathbf{A}_{\mathbf{tested}}$ lie in the space spanned by the columns of $\overline{\mathbf{U}}$ or the rows of $\overline{\mathbf{V}}$ respectively [32,45].

Equations for the Target Testing (TT) of $\mathbf{C}_{\mathbf{tested}}$ and $\mathbf{A}_{\mathbf{tested}}$ can be obtained by expressing Equations 2.26 and 2.27 as a function of $\mathbf{T}$, i.e. $\mathbf{T} = \overline{\mathbf{U}}^{\mathrm{T}}\mathbf{C}_{\mathbf{tested}}$ and $\mathbf{T}^{-1} = \mathbf{A}_{\mathbf{tested}}\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{S}}^{-1}$, and substituting back these respective expressions into Equations 2.26 and 2.27. Note, as $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ are orthonormal matrices and $\overline{\mathbf{S}}$ is a square matrix, $\overline{\mathbf{U}}^{+} = \overline{\mathbf{U}}^{\mathrm{T}}$ and $\left(\overline{\mathbf{SV}}\right)^{+} = \overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{S}}^{-1}$.

$$\mathbf{C}_{\mathbf{projected}} = \left(\overline{\mathbf{U}}\,\overline{\mathbf{U}}^{\mathrm{T}}\right)\mathbf{C}_{\mathbf{tested}} \tag{2.28}$$

$$\mathbf{A}_{\mathbf{projected}} = \mathbf{A}_{\mathbf{tested}}\left(\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{V}}\right) \tag{2.29}$$

Equations 2.28 and 2.29 allow target testing matrices $\mathbf{C}_{\mathbf{tested}}$ or $\mathbf{A}_{\mathbf{tested}}$ (called input target) by projecting them on the space spanned by matrices $\overline{\mathbf{U}}$ or $\overline{\mathbf{V}}$, resulting in projected matrices $\mathbf{C}_{\mathbf{projected}}$ or $\mathbf{A}_{\mathbf{projected}}$ (called output target), which ideally equal $\mathbf{C}_{\mathbf{tested}}$ or $\mathbf{A}_{\mathbf{tested}}$ [45]. The difference between the tested and the projected matrices is an indication of the goodness of the tested matrices that can be exploited to refine an initial guess for **C** (or **A**) and eventually obtain improved values for these matrices. This is the principle of Iterative Target Transformation Factor Analysis (ITTFA) [32,45,48].

$$\min_{\mathbf{C}} \left\| \mathbf{C} - \left(\overline{\mathbf{U}}\,\overline{\mathbf{U}}^{\mathrm{T}}\right)\mathbf{C} \right\|^2 \tag{2.30}$$

$$\min_{\mathbf{A}} \left\| \mathbf{A} - \mathbf{A}\left(\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{V}}\right) \right\|^2 \tag{2.31}$$

Note the similarity of Equations 2.30 and 2.31 obtained for ITTFA with Equations 2.21 (Furusjö et al) and 2.22 used in kinetic hard-modelling.

At the first iteration, the initial guess for **C** (or **A**) is directly used in Equation 2.30 (or 2.31). For the second to the last iterations, the input target **C** (or **A**) is iteratively replaced by its output target $\overline{\mathbf{U}}\,\overline{\mathbf{U}}^{\mathrm{T}}\mathbf{C}$ (or $\mathbf{A}\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{V}}$). At the last iteration, the linear counter part is finally determined by linear regression, i.e. $\mathbf{A} = \mathbf{C}^{+}\mathbf{Y}$ (or $\mathbf{C} = \mathbf{Y}\mathbf{A}^{+}$). The convergence of ITTFA towards the minimum of the objective function is known to be poor and sometimes no well defined minimum might even exist [32].

An initial guess for **C** and for its subsequent refinement by ITTFA can be obtained by Evolving Factor Analysis (EFA) under the form of a window of concentration (see 2.2.2.3). For the refinement of **A** by means of Equation 2.31, it can be difficult to obtain such an initial guess. That is why ITTFA on **A** is often not considered in practice. Note that soft constraints (e.g. non-negativity of **C**) are generally required at each iteration of ITTFA to reduce rotational ambiguity [45,48].

2.2.2.3 Evolving factor analysis

Classical Evolving Factor Analysis (EFA) is a method using SVD in a repeated way in order to identify the appearance and disappearance of species with time [45,48,49]. Instead of performing SVD on the entire matrix **Y** (*nt* x *nw*), EFA applies SVD on gradually growing sub-matrices by increasing the size of the previous matrix by one row. Thus, each sub-matrix is formed by the first *i* spectra of **Y**, where *i* varies from 1 to *nt*. In this forward direction, the appearance of new absorbing species is indicated by the emergence of a new significant singular value [32]. EFA can also be applied in a backward mode, applying SVD on sub-matrices comprised by the last *i* spectra of **Y**, *i* varying from 1 to *nt*. In this backward direction, the emergence of a significant singular value indicates the disappearance of an absorbing species. The significance level for the appearance/disappearance of singular values is fixed by statistical means or more commonly by visual inspection.

Performing forward and backward EFA allows defining a domain of existence, called window of concentration, for each reactive species [49]. This window can be used as initial guesses (for **C**) for iterative soft-modelling methods, such as ITTFA (see 2.2.2.2) or ALS (see 2.2.2.4). These concentration windows can also be used as initial guesses for the transformation matrix to be calculated by RFA (see 2.2.2.2), an initial matrix **T** being calculated by expressing Equation 2.26 as $\mathbf{T} = \bar{\mathbf{U}}^{+}\mathbf{C}$ [47].

Another variant of EFA has been introduced which consists in applying SVD on **Y** from top to bottom on sub-matrices of fixed dimensions. This method, called Fixed-Size Moving Window EFA (FSMW/EFA) contributes to the detection of selective zones and minor constituents [45,50].

2.2.2.4 Alternating least squares

Alternating Least Squares (ALS), also known as MCR-ALS, is a method for the refinement of an initial guess for **C** or **A** working on the principle of successive linear regressions [32,45,48]. However, rather than ITTFA, both matrices **C** and **A** are refined during the iterative process of ALS. When ALS is initialised with a guess for **C**, matrix **A** is calculated by linear regression ($\mathbf{A}=\mathbf{C}^{+}\mathbf{Y}$) and in turn is used to recalculate **C** by subsequent linear regression ($\mathbf{C}=\mathbf{Y}\mathbf{A}^{+}$). These iterative steps are repeated until convergence of $\|\mathbf{Y}-\mathbf{C}\mathbf{A}\|^2$. ALS can also be initialised with a guess for **A**. In such a case, **C** is calculated by $\mathbf{C}=\mathbf{Y}\mathbf{A}^{+}$ and is used to calculate $\mathbf{A}=\mathbf{C}^{+}\mathbf{Y}$, until a minimum is found in $\|\mathbf{Y}-\mathbf{C}\mathbf{A}\|^2$. Note that, soft-constraints are usually applied between each regression step in order to reduce rotational ambiguity between **C** and **A**.

2.2.3 Hybrid methods

Chemometric methods discussed in Sections 2.2.1 (hard-modelling) and 2.2.2 (soft-modelling) can be combined in order to join the strengths of the two approaches. These hybrid methods cover soft- combined with other soft-modelling methods (soft/soft), hard- allied to other hard-modelling approaches (hard/hard), and soft- combined with hard-modelling (soft/hard).

An example of soft/soft combination, already discussed in Section 2.2.2, is the use of EFA, i.e. of the resulting concentration window, as initial guess for ITTFA or ALS. An example of hard/hard hybrid methods would be the combination of implicit and explicit direct kinetic hard-modelling (**C** modelled) in spectroscopy [32,41] or in calorimetry [2]. This hybrid approach is discussed in details for spectroscopy in Chapter 4 (Strategy 2). The combination of soft- with hard-modelling, also referred to soft-modelling under hard constraints [51], leads certainly to the most interesting hybrid methods. In these hybrid methods, only a part of the matrix **Y**, for which a kinetic model is available, is fitted by hard-modelling (usually by direct implicit hard-modelling), the unknown part of **Y** being treated by soft-modelling [38,51]. Theoretically, this soft/hard approach allows optimising a part of the kinetic parameters only, without the need to elucidate the entire kinetic mechanism. In practice, however, combining soft- with hard-modelling is only applicable to some rare reactions, whose rate laws can be decoupled, or for the soft-modelling of some independent artefacts [51].

2.3 Appendix

2.3.1 Data structure and measurement order

Measured data can be classified according to their order (mode or way) that can be defined as the number of independent axis of resolution. Data order can be extended to higher ways by concatenating several individual data obtained under different experimental conditions. Data of *i*-th order (*i* being an integer greater or equal to zero) are measurements organised in an *i*-dimensional array and can be extended to (*i*+1)-way data [52]. Measurement orders up to three and related examples taken from spectroscopy are described in Table 2.1.

Table 2.1: Description of measurement orders from zero to three and examples taken from spectroscopy.

Order	Data structure [a)]	Example [b)]	Extension	Data structure [a)]	Example [c)]
0	Scalar	(1 x 1)	1-way	Vector	(1 x 1 x *ne*+1)
1	Vector	(1 x *nw*) (*nt* x 1)	2-way	Matrix	(1 x *nw* x *ne*+1) (*nt* x 1 x *ne*+1)
2	Matrix	(*nt* x *nw*) (nw_{ex} x nw_{em})	3-way [d)]	Array (3-dim)	(*nt* x *nw* x *ne*+1) (nw_{ex} x nw_{em} x *ne*+1)
3	Array (3-dim)	(*nt* x nw_{ex} x nw_{em})	4-way	Array (4-dim)	(*nt* x nw_{ex} x nw_{em} x *ne*+1)

a) In this context, an array is defined as a set of numbers organised in 3 or 4 dimensions. Note that scalars, vectors and matrices are respectively arrays of dimensions 0, 1 and 2.

b) Dimensions of data are given. For example, (1 x 1): absorbances at one time and one wavelength, (1 x *nw*): absorbance at one time and *nw* wavelengths, (nw_{ex} x nw_{em}): absorbances at nw_{ex} exciting wavelengths and nw_{em} emitting wavelengths.

c) Dimensions of data are given. For example, (1 x 1 x *ne*+1): absorbances at one time, one wavelength and (*ne*+1) different initial concentrations, (1 x *nw* x *ne*+1): absorbance at one time, *nw* wavelengths and (*ne*+1) different initial concentrations, (nw_{ex} x nw_{em} x *ne*+1): absorbances at nw_{ex} exciting wavelengths, nw_{em} emitting wavelengths and (*ne*+1) different initial concentrations.

d) Extension of second order data to 3-way data is also called second order global analysis.

2.3.2 Bilinearity, trilinearity and n-linearity

We define *n*-linear data as data of order *n* following a function of *n* variables that is linear in each variable. Most common values of *n* are two and three. For example, bilinear data are second order data following a function of two variables that is linear in each variable. Note that time and wavelength resolved absorbance data are bilinear data as they depend on a function of two linear variables, i.e. the concentrations and the absorptivities of the pure components.

$$y_{t,\lambda} = c_{t,1}a_{1,\lambda} + c_{t,2}a_{2,\lambda} + \ldots + c_{t,ns}a_{ns,\lambda} = \mathbf{c}_{t,:}\mathbf{a}_{:,\lambda} \tag{2.32}$$

Where t and λ are particular times and wavelengths respectively. Note that, in agreement with Matlab [53], the subscript : defines all elements along the corresponding dimension. Calorimetric data are also bilinear as they depend on a function of two variables, i.e. the derivatives of the reaction extents and the reaction enthalpies.

$$\dot{q}_{r,t} = -\dot{\xi}_{t,1}\Delta_r H_1 - \dot{\xi}_{t,2}\Delta_r H_2 - \ldots - \dot{\xi}_{t,nr}\Delta_r H_{nr} = \dot{\boldsymbol{\xi}}_{t,:}\left(-\boldsymbol{\Delta}_r\mathbf{H}_{:}\right) \tag{2.33}$$

Trilinear data could be obtained, for example, from a chemical reaction followed by repeated chromatography and hyphenated spectroscopy, i.e. at each reaction time t_r all chromatographic peaks are analysed by spectroscopy [25,54]. In this case, absorbance data depend on a function of three linear variables, i.e. the concentrations, the elution profiles and the absorptivities of the pure components.

$$y_{t_r,t_e,\lambda} = c_{t_r,1,1}e_{1,t_e,1}a_{1,1,\lambda} + c_{t_r,1,2}e_{1,t_e,2}a_{2,1,\lambda} + \ldots + c_{t_r,1,ns}e_{1,t_e,ns}a_{ns,1,\lambda} = \mathbf{c}_{t_r,1,:} \otimes \mathbf{e}_{1,t_e,:} \otimes \mathbf{a}_{:,1,\lambda} \tag{2.34}$$

With $\underline{\mathbf{Y}}$ a tensor (three dimensional array) of absorbance data of dimensions (nt_r x nt_e x nw), $\underline{\mathbf{C}}$ a tensor of concentration profiles of dimensions (nt_r x 1 x ns), $\underline{\mathbf{E}}$ a tensor of elution profiles of dimensions (1 x nt_e x ns) and $\underline{\mathbf{A}}$ a tensor of pure component spectra of dimensions (ns x 1 x nw). Tensor products are indicated by $\otimes$. Tensor dimensions are indicated with nt_r the number of reaction times (t_r being a particular reaction time) and nt_e the number of elution times (t_e being a particular elution time).

2.3.3 Definition of linear and non-linear parameters

Let g(**p**) be a function depending on a parameter vector **p** = [..., p_i, ..., p_j, ...]. A parameter p_i is defined as linear if the derivative of g with respect to p_i is not a function of p_i.

$$\frac{\mathrm{d}}{\mathrm{d}p_i}\mathrm{g}(\mathbf{p}) \neq \mathrm{g}'(p_i) \tag{2.35}$$

A parameter p_j is defined as non-linear if the derivative of g with respect to p_j is still a function of p_j.

$$\frac{\mathrm{d}}{\mathrm{d}p_j}\mathrm{g}(\mathbf{p}) = \mathrm{g}'(p_j) \tag{2.36}$$

Except for zero order rate laws, rate constants are always non-linear parameters with respect to the concentration matrix **C** and to any type of objective function defined in Chapter 2.

2.3.4 Left and right pseudo-inverse of a matrix

In spectroscopy, and according to Beer's law, concentration profiles **C** (*nt* x *ns*) and pure component spectra **A** (*ns* x *nw*) are linear parameters with respect to the spectroscopic data matrix **Y** (*nt* x *nw*).

In case **C** and **Y** are given, **A** can be calculated by linearly regressing **Y** on **C**. An expression for this linear regression can be obtained by multiplying Beer's law on both sides by the left pseudo-inverse of **C**. In Matlab [53], this operation is performed by using the back slash operator.

$$\mathbf{A} = \mathbf{C}^{+}\mathbf{Y} = \mathbf{C} \setminus \mathbf{Y} \qquad \text{with } \mathbf{C}^{+} = \left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}} \tag{2.37}$$

In case **A** and **Y** are given, **C** can be calculated by linearly regressing **Y** on **A**. An expression for this linear regression can be obtained by multiplying Beer's law on both sides by the right pseudo-inverse of **A**. In Matlab, this operation is performed by using the forward slash operator.

$$\mathbf{C} = \mathbf{Y}\mathbf{A}^{+} = \mathbf{Y} / \mathbf{A} \qquad \text{with } \mathbf{A}^{+} = \mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\,\mathbf{A}^{\mathrm{T}}\right)^{-1} \tag{2.38}$$

In calorimetry, the derivative of the reaction extents $\dot{\boldsymbol{\xi}}$ (*nt* x *nr*) and the reaction enthalpies $\boldsymbol{\Delta}_r\mathbf{H}$ (*nr* x 1) are linear parameters with respect to the calorimetric data vector $\dot{\mathbf{q}}_r$ (*nt* x 1).

When $\dot{\boldsymbol{\xi}}$ and $\dot{\mathbf{q}}_r$ are given, reaction enthalpies can be calculated by linearly regressing $\dot{\mathbf{q}}_r$ on $\dot{\boldsymbol{\xi}}$. An expression for this linear regression can be obtained by multiplying the energy balance on both sides by the left pseudo-inverse of $\dot{\boldsymbol{\xi}}$ or using the back slash operator of Matlab.

$$\boldsymbol{\Delta}_r\mathbf{H} = -\dot{\boldsymbol{\xi}}^{+}\dot{\mathbf{q}}_r = -\dot{\boldsymbol{\xi}} \setminus \dot{\mathbf{q}}_r \qquad \text{with } \dot{\boldsymbol{\xi}}^{+} = \left(\dot{\boldsymbol{\xi}}^{\mathrm{T}}\dot{\boldsymbol{\xi}}\right)^{-1}\dot{\boldsymbol{\xi}}^{\mathrm{T}} \tag{2.39}$$

When $\boldsymbol{\Delta}_r\mathbf{H}$ and $\dot{\mathbf{q}}_r$ are known, the extents of reaction can be calculated by linearly regressing $\dot{\mathbf{q}}_r$ on $\boldsymbol{\Delta}_r\mathbf{H}$. An expression for this linear regression can be obtained by multiplying the energy balance on both sides by the right pseudo-inverse of $\boldsymbol{\Delta}_r\mathbf{H}$ or using the forward slash operator of Matlab.

$$\dot{\boldsymbol{\xi}} = -\dot{\mathbf{q}}_r\left(\boldsymbol{\Delta}_r\mathbf{H}\right)^{+} = -\dot{\mathbf{q}}_r / \boldsymbol{\Delta}_r\mathbf{H} \qquad \text{with } \left(\boldsymbol{\Delta}_r\mathbf{H}\right)^{+} = \boldsymbol{\Delta}_r\mathbf{H}^{\mathrm{T}}\left[\boldsymbol{\Delta}_r\mathbf{H}\left(\boldsymbol{\Delta}_r\mathbf{H}^{\mathrm{T}}\right)\right]^{-1} \tag{2.40}$$

2.4 References

[1] A. Zogg, F. Stoessel, U. Fischer, K. Hungerbühler, Thermochim. Acta, 419 (2004), 1-17.
[2] G. Puxty, M. Maeder, R.R. Rhinehart, S. Alam, S. Moore, P.J. Gemperline, J. Chemom., 19 (2005), 329-340.
[3] W. Regenass, J. Therm. Anal., 49 (1997), 1661-1675.
[4] J. Singh, Proc. Safety Prog., 16 (1997), 43-49.
[5] O. Ubrich, B. Srinivasan, P. Lerena, D. Bonvin, F. Stoessel, Chem. Eng. Sci., 56 (2001), 5147-5156.
[6] F. Stoessel, O. Ubrich, J. Therm. Anal. Cal., 64 (2001), 61-74.
[7] K.R. Westerterp, E.J. Molga, Chemical Engineering Research & Design, 84 (2006), 543-552.
[8] A. Zogg, M. Wohlwend, U. Fischer, K. Hungerbühler, European Patent EP 1 184 649 A1, 2000.
[9] F. Visentin, S.I. Gianoli, A. Zogg, O.M. Kut, K. Hungerbühler, Org. Proc. Res. Dev., 8 (2004), 725-737.
[10] J. Pastré, Doctoral Thesis 13840, ETH Zürich, Switzerland, 2000.
[11] A. Zogg, Doctoral Thesis 15086, ETH Zürich, Switzerland, 2003.
[12] G. Richner, Doctoral Thesis 17972, ETH Zürich, Switzerland, 2008.
[13] F. Visentin, Doctoral Thesis 16053, ETH Zürich, Switzerland, 2005.
[14] J. Pastre, A. Zogg, U. Fischer, K. Hungerbühler, Org. Proc. Res. Dev., 5 (2001), 158-166.
[15] A.E. Croce, J. Chem. Educ., 79 (2002), 506-509.
[16] G. Puxty, U. Fischer, M. Jecklin, K. Hungerbühler, Chimia, 60 (2006), 605-610.
[17] S.I. Gianoli, G. Puxty, U. Fisher, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 85 (2007), 47-62.
[18] F. Visentin, G. Puxty, O.M. Kut, K. Hungerbühler, Ind. Eng. Chem. Res., 45 (2006), 4544-4553.
[19] D.A. Skoog, D.M. West, F.J. Holler, Fundamentals of analytical chemistry, 7th Edition, Saunders College Publishing, Philadelphia, USA, 1991.
[20] A.C. Quinn, P.J. Gemperline, B. Baker, M. Zhu, D.S. Walker, Chemom. Intell. Lab. Syst., 45 (1999), 199-214.
[21] E.N.M. van Sprang, H.J. Ramaker, H.F.M. Boelens, J.A. Westerhuis, D. Whiteman, D. Baines, I. Weaver, Analyst, 128 (2003), 98-102.
[22] A. Bayada, G.A. Lawrance, M. Maeder, K.J. Molloy, Appl. Spectrosc., 49 (1995), 1789-1792.
[23] R.M. Silverstein, G.C. Basler, T.C. Morill, Spectrometric identification of organic compounds, Fifth Edition, Wiley, New-York, USA, 1991.
[24] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Second Edition, Wiley, Weinheim, Germany, 2007.
[25] R.G. Brereton, Chemometrics - Data Analysis for the Laboratory and Chemical Plant, Wiley, England, 2003.
[26] P. Gemperline (Ed.), Practical Guide to Chemometrics, Taylor and Francis, Boca Raton, USA, 2006.
[27] E.R. Malinowski, Factor Analysis in Chemistry, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
[28] G. Puxty, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 81 (2006), 149-164.

[29] E. Kriesten, D. Mayer, F. Alsmeyer, C.B. Minnich, L. Greiner, W. Marquardt, Chemom. Intell. Lab. Syst., 93 (2008), 108-119.
[30] V.M. Taavitsainen, H. Haario, J. Chemom., 15 (2001), 215-239.
[31] V.M. Taavitsainen, H. Haario, M. Laine, J. Chemom., 17 (2003), 140-150.
[32] M. Maeder, Y.M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, NL, 2007.
[33] P. Geladi, B.R. Kowalski, Anal. Chim. Acta, 185 (1986), 1-17.
[34] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, Comput. Chem. Eng., 20 (1996), 1451-1465.
[35] P. Gemperline, G. Puxty, M. Maeder, D. Walker, F. Tarczynski, M. Bosserman, Anal. Chem., 76 (2004), 2575-2582.
[36] M. Maeder, A.D. Zuberbühler, Anal. Chem., 62 (1990), 2220-2224.
[37] M. Maeder, Y.M. Neuhold, Chapter 7 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 218-256.
[38] H. Haario, V.M. Taavitsainen, Chemom. Intell. Lab. Syst., 44 (1998), 77-98.
[39] J.E. Carlson, V.M. Taavitsainen, J. Chemom., 22 (2008), 752-757.
[40] E. Furusjö, L.G. Danielsson, Chemom. Intell. Lab. Syst., 50 (2000), 63-73.
[41] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta, 396 (1999), 231-240.
[42] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C++ - The art of Scientific Computing, Second Edition, Cambridge University Press, New York, USA, 2005.
[43] G. Puxty, University of Newcastle, Newcastle, Australia, 2004.
[44] J.H. Jiang, Y. Ozaki, Applied Spectroscopy Reviews, 37 (2002), 321-345.
[45] A. de Juan, E. Casassas, R. Tauler, in Encyclopedia of Analytical Chemistry, Meyers, R. A. (Ed.), Wiley, Chichester, England, 2000, 9800-9837.
[46] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, M. Maeder, J. Chemom., 20 (2006), 302-310.
[47] C. Mason, M. Maeder, A. Whitson, Anal. Chem., 73 (2001), 1587-1594.
[48] R. Tauler, A. de Juan, Chapter 11 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 417-474.
[49] M. Maeder, Anal. Chem., 59 (1987), 527-530.
[50] H.R. Keller, D.L. Massart, Anal. Chim. Acta, 246 (1991), 379-390.
[51] A. de Juan, M. Maeder, M. Martinez, R. Tauler, Chemom. Intell. Lab. Syst., 54 (2000), 123-141.
[52] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, Trac-Trends Anal. Chem., 26 (2007), 752-765.
[53] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007; http://www.mathworks.com
[54] Y.M. Neuhold, M. Maeder, J. Chemom., 16 (2002), 218-227.

CHAPTER 3

Uncertainties and error propagation in kinetic hard-modelling of spectroscopic data

This chapter was published as an article under the reference:

J. Billeter, Y.-M. Neuhold, L. L. Simon, G. Puxty, K. Hungerbühler
Chemometrics and Intelligent Laboratory Systems, 93 (2008), 120-131
http://dx.doi.org/10.1016/j.chemolab.2008.05.001

* * *

3.1 Abstract

A novel method is presented for the rigorous propagation of uncertainties in initial concentrations and in dosing rates into the errors in the rate constants fitted by multivariate kinetic hard-modelling of spectroscopic data using the Newton-Gauss Levenberg/Marquardt optimisation algorithm. The method was successfully validated by Monte-Carlo sampling. The impact of the uncertainties in initial concentrations and in the dosing rate was quantified for simulated spectroscopic data based on a second and a formal third order rate law under batch and semi-batch conditions respectively. An important consequence of this study regarding optimum experimental design is the fact that the propagated error in a second order rate constant is minimal under exact stoichiometric conditions or when the reactant with the lowest associated uncertainty in its initial concentration is in a reasonable excess (pseudo first order conditions). As an experimental example, the reaction of benzophenone with phenylhydrazine in THF was investigated repeatedly (17 individual experiments) by UV-vis and mid-IR spectroscopy under the same semi-batch conditions, dosing the catalyst acetic acid. For all experiments and spectroscopic signals, reproducible formal third order rate constants were determined.

Applying the proposed method of error propagation to any single experiment, it was possible to predict 80% (UV-vis) and 40% (mid-IR) of the observed standard deviation in the rate constants obtained from all experiments. The largest contribution to this predicted error in the rate constant could be assigned to the dosing rate. The proposed method of error propagation is flexible and can straightforwardly be extended to propagate other possible sources of error.

3.2 Introduction

It is a common trend in the chemical and pharmaceutical industry to maximise yield and safety while minimising the waste produced during manufacturing [1-3]. New processes need to be designed accordingly and existing ones may require re-evaluation of the process conditions in order to meet new standards [4].

There are various techniques in order to optimise a reaction process based on spectroscopic measurements. Soft-modelling methods such as principal component analysis [5], evolving factor analysis [6] and alternating least squares [7-10] have no or limited predictive capability. Predictive capability is possible for calibration methods [11] such as principal component regression [5], partial least squares [12] and neural networks [13]. However, these methods only allow for interpolation within the calibration set and are not suitable for extrapolation. The method used in this work, kinetic hard-modelling, does allow for both interpolation and extrapolation when an appropriate empirical or molecular rate law is applied reflecting the underlying reaction mechanism [14-19]. This task is facilitated by the recent development of multivariate analytical in-situ devices and optimisation algorithms [20,21].

For kinetic analysis via absorption spectroscopy (e.g. UV-vis and mid-IR), two classes of model parameters can be distinguished based on their relationship to the measured signal. Rate constants that define the kinetic model and thus the concentration profiles are non-linear parameters with respect to the measured absorbance. According to Beer's law, molar component spectra are linear parameters that can be eliminated from the optimisation problem [14]. It is then the task of optimisation algorithms to minimise the difference between the experimental and modelled absorbance data.

In recent years, the Newton-Gauss Levenberg/Marquardt (NGL/M) algorithm that minimises this difference in the least squares sense has become the method of choice to solve such problems [14,19,22-25].

As a gradient method, the NGL/M algorithm directly allows estimation of the uncertainty in the fitted rate constants based on the corresponding variance/covariance matrix from one measurement only [24,25]. However, this calculation systematically underestimates this uncertainty when compared to the one resulting from multiple repetitions of the experiment under the same conditions.

Some other contributions to the uncertainty in the rate constants, such as baseline shifts [26,27], spectral constraints in the least squares [28] and preceding calibration procedures [29] have been studied. Also, bootstrapping has been compared to variance-covariance based uncertainty calculations [30]. However, under strictly controlled isothermal conditions, and provided the kinetic model is correct, uncertainties in the initial concentrations and/or dosing rates (semi-batch) are also likely to represent major contributing factors. To our knowledge, these uncertainties have not yet been incorporated into a kinetic hard-modelling procedure.

In the present article, the impact of these additional uncertainties is investigated and quantified by classical error propagation. In this context, based on some selected kinetic mechanisms, simulated spectroscopic data are analysed under batch or semi-batch conditions by the NGL/M algorithm combined with error propagation. This method was also applied to the reaction of benzophenone with phenylhydrazine [31] and the predicted error in the rate constant was compared to the uncertainty obtained from multiple experimental repetitions. The reaction was followed by UV-vis and mid-IR spectroscopy under strictly isothermal semi-batch conditions.

3.3 Theoretical considerations

Kinetic modelling applied to spectroscopy relies on Beer's law to decompose a measured signal into the concentration and the molar spectra of the pure components. Beer's law can be written elegantly in matrix notation.

$$\mathbf{Y} = \mathbf{CA} + \mathbf{R} \tag{3.1}$$

Where **Y** (*nt* x *nw*) represents the spectroscopic measurements, i.e. the time and wavelength/wavenumber resolved absorbance signals, **C** (*nt* x *ns*) the concentration profiles of the contributing species and **A** (*ns* x *nw*) the pure component spectra, i.e. the molar absorptivities multiplied by the path length. Here, *nt* is the number of reaction times, *nw* the number of wavelengths/wavenumbers and *ns* the total number of species. The residual matrix **R** (*nt* x *nw*) comprises the deviation from Beer's law due to inherent experimental noise. It is assumed that the baseline does not change with time.

The matrix of concentration profiles **C** is calculated by numerical integration of the system of ordinary differential equations describing the kinetic hard model by elementary steps of the corresponding chemical equations [17,32]. Note that **C** is a function of the selected model, the rate constants **k** (1 x *nr*), the initial concentrations $\mathbf{c_0}$ (1 x *ns*) and the dosing (flow) rate of each dosing step, $\mathbf{f_s}$ (1 x *nf*); *nr* denotes the number of reaction steps (one rate constant for each step), *nf* is the number of different dosing steps. In kinetic hard-modelling, $\mathbf{c_0}$ and **f** are treated as known a priori information and generally not fitted.

$$\mathbf{R}(\text{model}, \mathbf{k}, \mathbf{c_0}, \mathbf{f_s}) = \mathbf{Y} - \mathbf{C}(\text{model}, \mathbf{k}, \mathbf{c_0}, \mathbf{f_s})\mathbf{A} \tag{3.2}$$

In least-squares analysis, the sum of squares, *ssq*, is calculated from the residuals **R** and used as the objective function to be minimised by iteratively changing **k**. The gradient-based Newton-Gauss Levenberg/Marquardt algorithm (NGL/M) [14,19,22-25] is used to solve this non-linear regression.

$$ssq = \sum_{i=1}^{nt}\sum_{j=1}^{nw} r_{i,j}^2 \tag{3.3}$$

Note that **A** is comprised by linear parameters only and can be eliminated from the non-linear optimisation by its linear least-squares estimate according to Equation 3.4 [14].

$$\mathbf{A} = \mathbf{C}^{+}\mathbf{Y} = \left(\mathbf{C}^{T}\mathbf{C}\right)^{-1}\mathbf{C}^{T}\mathbf{Y} \tag{3.4}$$

where $\mathbf{C}^{+}$ denotes the pseudo-inverse of **C**. Importantly, baseline variations along the wavelengths/wavenumbers have no impact on the least squares minimum, they only affect the fitted pure component spectra **A**.

Any gradient-based optimisation method requires the calculation of a Jacobian, i.e. the first partial derivative of the residuals **R** with respect to **k**. As this would lead to a three dimensional Jacobian, it is convenient to unfold **R** into a long vector **r** (*nt*·*nw* x 1) [18]. Then, the Jacobian **J** (*nt*·*nw* x *nr*) can be calculated by a forward finite difference, as illustrated by Equation 3.5.

$$\mathbf{J} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{k}}\right) = \left[\left(\frac{\partial \mathbf{r}}{\partial k_1}\right) \quad \left(\frac{\partial \mathbf{r}}{\partial k_2}\right) \quad \dots \quad \left(\frac{\partial \mathbf{r}}{\partial k_{nr}}\right)\right] \text{ with } \left(\frac{\partial \mathbf{r}}{\partial k_i}\right) \approx \frac{\mathbf{r}\left(\mathbf{k} + \delta k_i\right) - \mathbf{r}\left(\mathbf{k}\right)}{\delta k_i} \tag{3.5}$$

where k_i is the *i*-th rate constant and δk_i the finite difference applied to the *i*-th rate constant.

The Jacobian is used by the NGL/M algorithm to iteratively shift **k** towards an optimum. For details see [17]. **J** is also used to approximate the Hessian **H** (*nr* x *nr*) ≈ $\mathbf{J}^{T}\mathbf{J}$ [19]. Based on the inverted Hessian $\mathbf{H}^{-1}$, the variance $\sigma_{k_i}^2$ associated with the *i*-th rate constant can be estimated from the variance/covariance matrix.

$$\sigma_{k_i}^2 = d_{i,i}\sigma_{\mathbf{r}}^2 \tag{3.6}$$

Where $d_{i,i}$ denotes the *i*-th diagonal element of the inverted Hessian $\mathbf{H}^{-1}$.

The scalar $\sigma_{\mathbf{r}}^2$, the variance of the residuals, is an estimate of the true variance $\sigma_{\mathbf{y}}^2$ in the measurement matrix **Y**, provided the noise is normally distributed and homoscedastic. It can be calculated from Equation 3.7.

$$\sigma_{\mathbf{r}}^2 = \frac{ssq}{nt \cdot nw - \left(nr + ns \cdot nw\right)} \approx \sigma_{\mathbf{y}}^2 \tag{3.7}$$

The denominator of this equation represents the degrees of freedom, i.e. the number of experimental data points (*nt*·*nw*) minus the number of fitted rate constants (*nr*) and fitted molar spectra of all absorbing species (*ns*·*nw*).

Equation 3.6 can be written in convenient matrix notation.

$$\boldsymbol{\sigma}_{\mathbf{k}}^2 = \mathrm{diag}\left(\mathbf{H}^{-1}\right)\sigma_{\mathbf{r}}^2 \tag{3.8}$$

Note that, in analogy to Matlab [33], the diag operator extracts a vector of diagonal elements from the corresponding matrix. The iterative process performed by the NGL/M algorithm in order to optimise the vector of rate constants **k** is outlined in Figure 3.1.

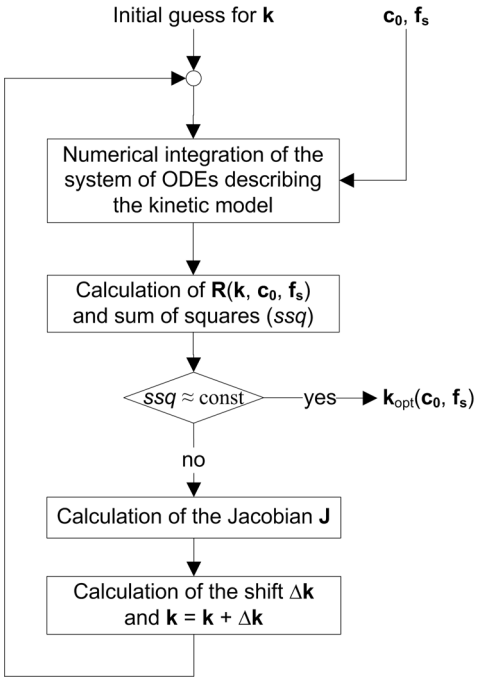

Figure 3.1: Simplified scheme of the NGL/M algorithm used in kinetic hard-modelling.

The vector of variances, $\boldsymbol{\sigma}_{\mathbf{k}}^2$, calculated by Equation 3.8, only covers the variances due to the residuals and the sensitivities of the rate constants with respect to these residuals. However, other additional sources of error are inherently present. Amongst these, the uncertainties in the initial concentrations and in the dosing rates are likely to have a significant impact on the variances of the calculated rate constants, provided the temperature is constant during the reaction and its fluctuation can be neglected. Equation 3.6 and 3.8 can be adapted to consider these additional uncertainties by classical error propagation [34].

$$\sigma_{k_\mathrm{i}}^2 = d_{\mathrm{i,i}}\sigma_{\mathbf{r}}^2 + \sum_{n=1}^{ns}\left(\frac{\partial k_\mathrm{i}}{\partial c_{0,\mathrm{n}}}\right)^2 \sigma_{c_{0,\mathrm{n}}}^2 + \sum_{m=1}^{nf}\left(\frac{\partial k_\mathrm{i}}{\partial f_{s,\mathrm{m}}}\right)^2 \sigma_{f_{s,\mathrm{m}}}^2 \tag{3.9}$$

Or in matrix notation:

$$\boldsymbol{\sigma}_{\mathbf{k}}^2 = \underbrace{\mathrm{diag}\left(\mathbf{H}^{-1}\right)\sigma_{\mathbf{r}}^2}_{=\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2} + \underbrace{\mathrm{diag}\left[\left(\frac{\partial \mathbf{k}}{\partial \mathbf{c_0}}\right)^{\mathrm{T}} \mathrm{DIAG}\left(\boldsymbol{\sigma}_{\mathbf{c_0}}^2\right)\left(\frac{\partial \mathbf{k}}{\partial \mathbf{c_0}}\right)\right]}_{=\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2} + \underbrace{\mathrm{diag}\left[\left(\frac{\partial \mathbf{k}}{\partial \mathbf{f_s}}\right)^{\mathrm{T}} \mathrm{DIAG}\left(\boldsymbol{\sigma}_{\mathbf{f_s}}^2\right)\left(\frac{\partial \mathbf{k}}{\partial \mathbf{f_s}}\right)\right]}_{=\boldsymbol{\sigma}_{\mathbf{k},\mathbf{f_s}}^2} \tag{3.10}$$

In Equation 3.10, vectors $\boldsymbol{\sigma}_{\mathbf{c_0}}^2$ and $\boldsymbol{\sigma}_{\mathbf{f_s}}^2$ contain the variances corresponding to the initial concentrations $\mathbf{c_0}$ and to the dosing rates $\mathbf{f_s}$. The partial derivatives $\partial\mathbf{k}/\partial\mathbf{c_0}$ and $\partial\mathbf{k}/\partial\mathbf{f_s}$ comprise the sensitivities of the rate constants with respect to the initial concentrations and with respect to the dosing rates. Throughout this manuscript, we will refer to the first term of Equation 3.10 as the variance of $\mathbf{k}$ due to the residuals, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2$, to the second term as the variance due to the initial concentrations, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$, and to the last term as the variance due to the dosing rates, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{f_s}}^2$.

Note that the upper case DIAG operator generates a diagonal matrix from the corresponding vector argument and thus performs the reverse operation compared to the lower case diag operator introduced previously.

The sensitivity factors $\partial\mathbf{k}/\partial\mathbf{c_0}$ and $\partial\mathbf{k}/\partial\mathbf{f_s}$ were evaluated numerically and estimated by the method of central finite differences. The use of this method, as opposed to the forward finite difference, was required to improve the numerical stability of the derivatives.

$$\left(\frac{\partial\mathbf{k}}{\partial c_{0,\mathrm{n}}}\right)\approx\frac{\mathbf{k}_{\mathrm{opt}}\left(\mathbf{c_0}+\delta c_{0,\mathrm{n}},\,\mathbf{f_s}\right)-\mathbf{k}_{\mathrm{opt}}\left(\mathbf{c_0}-\delta c_{0,\mathrm{n}},\,\mathbf{f_s}\right)}{2\delta c_{0,\mathrm{n}}} \tag{3.11}$$

$$\left(\frac{\partial\mathbf{k}}{\partial f_{\mathrm{s,m}}}\right)\approx\frac{\mathbf{k}_{\mathrm{opt}}\left(\mathbf{c_0},\,\mathbf{f_s}+\delta f_{s,\mathrm{m}}\right)-\mathbf{k}_{\mathrm{opt}}\left(\mathbf{c_0},\,\mathbf{f_s}-\delta f_{s,\mathrm{m}}\right)}{2\delta f_{s,\mathrm{m}}} \tag{3.12}$$

The finite differences $\delta c_{0,\mathrm{n}}$ and $\delta f_{s,\mathrm{m}}$ used to calculate the derivatives were set to 0.1% of the corresponding values $c_{0,\mathrm{n}}$ and $f_{s,\mathrm{m}}$. This ensured the numerical stability for all examples presented in this article. $\mathbf{k_{opt}}$ refers to the calculated optimum rate constants that minimise the residuals in the least-squares sense. For each finite difference by which an initial concentration or the dosing rate is modified, $\mathbf{k_{opt}}$ has to be re-determined by the NGL/M algorithm. The whole procedure is presented in Figure 3.2.

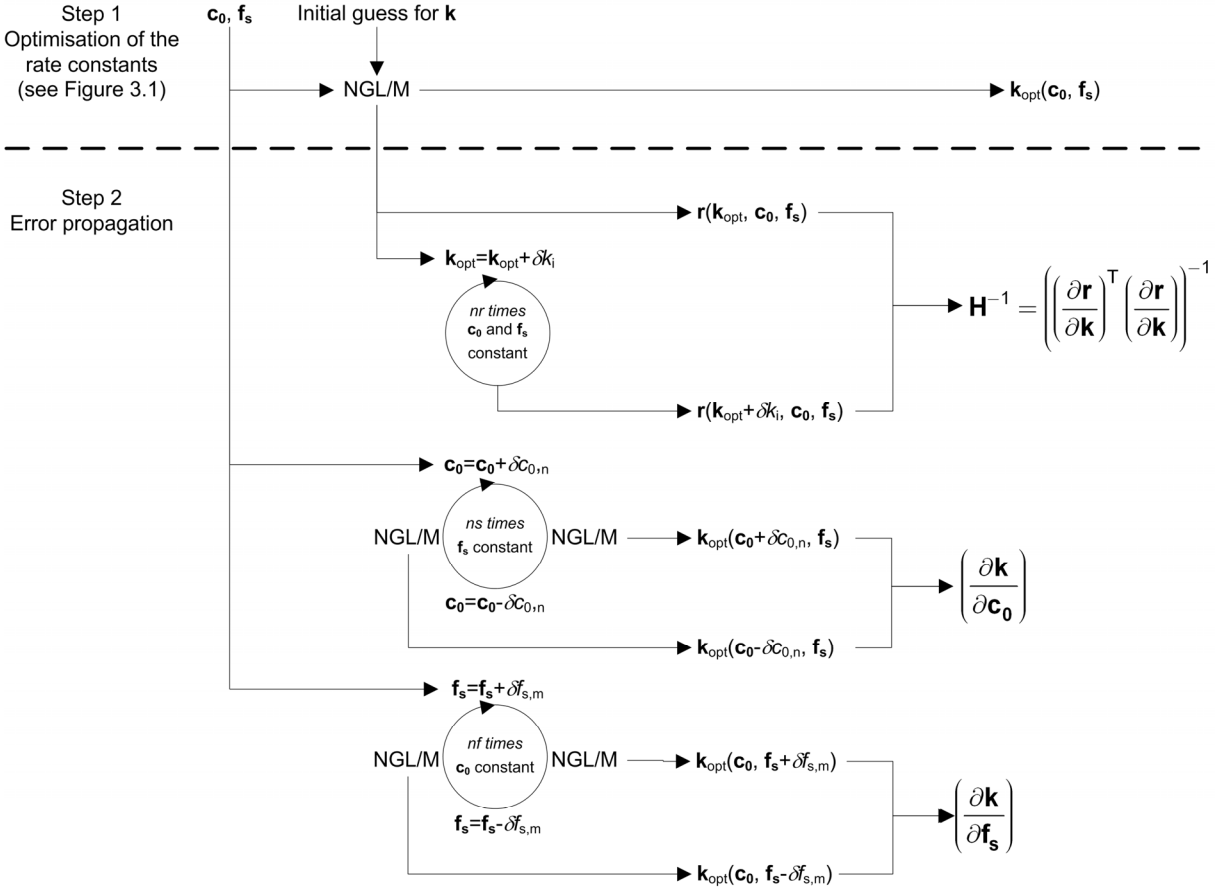

Figure 3.2: Simplified scheme to determine the inverted Hessian $\mathbf{H}^{-1}$ and the derivatives $\partial\mathbf{k}/\partial\mathbf{c_0}$ and $\partial\mathbf{k}/\partial\mathbf{f_s}$ calculated by the NGL/M algorithm.

As indicated by Figure 3.1 and Figure 3.2, error propagation problems of this kind cannot be solved explicitly. This is due to the fact that outer non-linear regression coupled to the inner nested integration of rate laws has no explicit solution except for some rare cases. Therefore an iterative computation is required.

3.4 Simulations and experiments

3.4.1 Simulations

Additional contributions to the variance in the rate constant according to Equation 3.10 are first investigated with simulated mechanisms. The simulated kinetic schemes considered here are summarised in Table 3.1.

Table 3.1: List of the mechanisms used to generate simulated data.

Mechanism	Rate law [a]	**k** [b]	Order	Dosing	Section
$A + B \rightarrow P$ or S [c]	$dc_{t,P}/dt = dc_{t,S}/dt = k \cdot c_{t,A} c_{t,B}$	0.5	2	No	3.5.1.1
$A + B + C \rightarrow P + C$ with C as catalyst	$dc_{t,P}/dt = k \cdot c_{t,A} c_{t,B} c_{t,C} - \frac{1}{v_t}\left(\frac{dv_t}{dt}\right) c_{t,P}$ [d] $dc_{t,C}/dt = \frac{1}{v_t}\left(\frac{dv_t}{dt}\right)\left(c_{in,C} - c_{t,C}\right)$	$1.75 \cdot 10^{-4}$	3 [e]	Yes	3.5.1.2

[a] The remaining derivatives can be calculated by closure
[b] $[L \cdot mol^{-1} \cdot s^{-1}]$ for 2nd and $[L^2 \cdot mol^{-2} \cdot s^{-1}]$ for 3rd order rate constants
[c] Pure component spectra of P and S were generated to produce different overlaps with A and B (see Figure 3.3)
[d] The change of volume per unit of time, dv_t/dt, is the flow rate f_s associated with the dosing event for species C of dosed catalyst concentration $c_{in,C}$
[e] Observed order due to a steady state assumption (see Section 3.5.1.2 for details)

Note that we defer from an investigation of zero and first order mechanisms. Formally, zero order kinetics depend on the initial concentration. However, as the rate is constant, the rate constant is linear with respect to the concentration profiles **C**. Thus it cannot be distinguished (and separated) from the molar spectra (that are also linear with **C**) within the fitting process without a priori knowledge on the pure component spectrum of the product. Incorporating a pure spectrum into the NGL/M fitting would require the propagation of its associated uncertainty. This was beyond the scope of this article.

Naturally, first order kinetics are independent from initial concentrations, i.e. the optimum of the NGL/M fitting is invariant to their change. Therefore there is no associated error propagation for first order rate constants.

Simulations were performed using Matlab [33]. Based on the selected molecular mechanism, the initial concentrations and the specified dosing rate, the corresponding system of differential equations was integrated by a 4^{th} order Runge-Kutta method (Matlab's ode45 solver) resulting in the concentration profiles **C** for the individual species. Pure component spectra **A** were generated using Gaussian functions, displayed in Figure 3.3. According to equation 3.1, spectroscopic absorbance data **Y** were generated by multiplication of **C** and **A**. Normally distributed noise with a constant absolute standard deviation σ_y = 10^{-4} was added, accounting for 0.01–0.02% relative to the maximum absorbance of the corresponding **Y**.

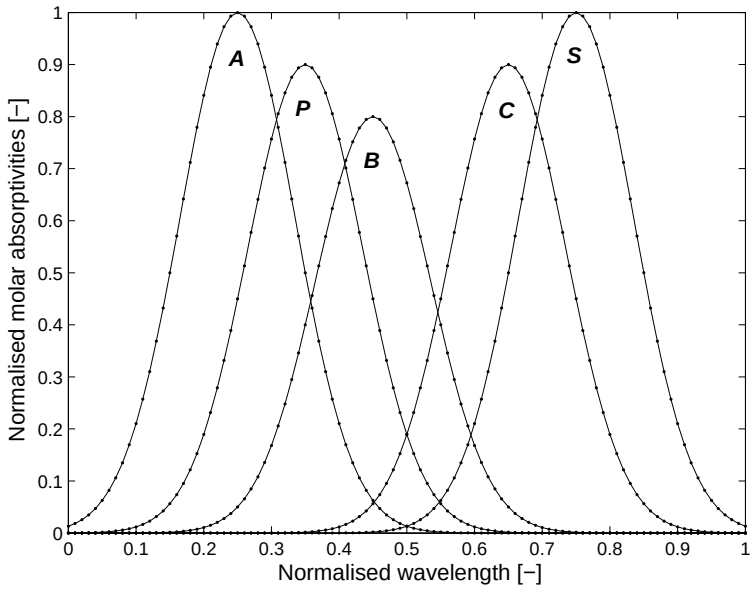

Figure 3.3: Pure component spectra as a function of wavelength. The spectra were generated using Gaussian functions centred at 0.25 (species *A*), 0.35 (*P*), 0.45 (*B*), 0.65 (*C*), 0.75 (*S*) with a constant half width of 0.2. Species names correspond to the models given in Table 3.1. For the sake of simplicity, component spectra are set to a maximum of one and wavelengths are evenly distributed between 0 and 1.

For the second order mechanism in Table 3.1, data were simulated for ten thousand data points up to an end time representing 95% conversion. The third order model was chosen to correspond to the mechanism, time range, initial conditions, dosing rate and associated uncertainties that were the basis for the analysis of the experimental data. The mean optimum rate constant as determined from the fitting of all experimental data (see Section 3.4.2) was used for the simulation ($1.75 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$). Pure component spectra were used according to Figure 3.3.

Using the third order model, the validation of the proposed method for error propagation was done by Monte-Carlo sampling. For this, a set of 10^4 normally distributed initial concentrations and dosing rates was generated. The means were chosen to be the same as the experimental initial concentrations and dosing rate; the associated standard deviations were taken from a preceding error estimation covering the experimental preparation procedure (see Section 3.4.2.1 and Section 3.7.1 – Appendix). For each of the Monte-Carlo samples the rate constant was re-optimised. The standard deviation of these 10^4 calculated rate constants was then compared to its prediction from the propagation of the error in the mean initial concentrations and dosing rate.

3.4.2 Experiments

The reaction of benzophenone with phenylhydrazine under acidic excess can be followed both in the mid-IR and UV-vis spectral range [31]. The reaction scheme is given in Figure 3.4.

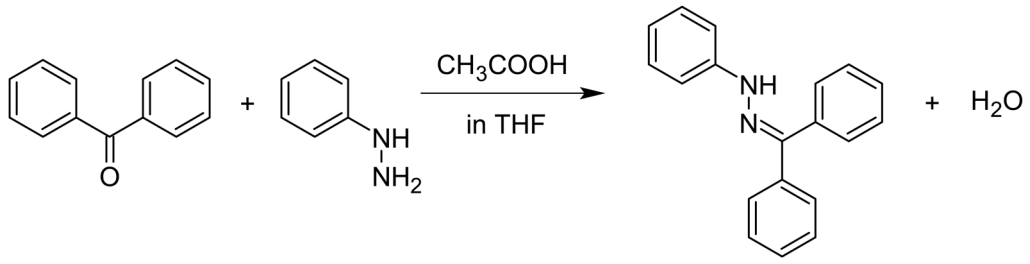

Figure 3.4: Reaction between benzophenone and phenylhydrazine catalysed by acetic acid to form benzophenone phenylhydrazone and water.

3.4.2.1 Sample preparation

Seventeen solutions of benzophenone (Fluka purum, certified 99.9%) in THF (Acros Organics for analysis) were prepared independently by weighing between 7.3022 g and 7.3027 g of benzophenone into a volumetric flask and making up to 50 mL with THF. Phenylhydrazine (Aldrich-Fine Chemicals, certified 99.6%) solutions were prepared by weighing between 13.0004 g and 13.0010 g of phenylhydrazine into a 50 mL flask and making up to 50 mL with THF. At the start, the reactor vessel was charged with 15 mL of each solution and thermostated at 25°C. This resulted in initial concentrations between 0.40033 $mol \cdot L^{-1}$ and 0.40036 $mol \cdot L^{-1}$ of benzophenone, and between 1.19737 $mol \cdot L^{-1}$ and 1.19743 $mol \cdot L^{-1}$ of phenylhydrazine. Note that for calculation of all concentrations, manufacturer's specifications for the purity were considered.

Error estimation in these initial concentrations due to sample preparation was performed by considering the following factors: standard deviation due to weighing (± 0.0001 g), due to filling the volumetric flasks (± 0.06 mL), and due to pipetting (± 0.04 mL/pipette). For details, see Section 3.7.1 (Appendix). As the differences in weighed samples are very small, the impact of weighing is negligible compared to filling the flask and pipetting. This leads to one common set of initial concentrations with an associated uncertainty of 0.292% for both compounds and all 17 experiments. Note that manufacturer's uncertainties in the stated purity of the reactants were not available and so were not propagated through the sample preparation.

Initially, the dosing pump was calibrated at 8.17 $mL \cdot min^{-1}$ by repeatedly weighing delivered volumes of water at 25°C. The corresponding standard deviation (0.14 $mL \cdot min^{-1}$) was close to the manufacturer's specification (0.2 $mL \cdot min^{-1}$). For the dosed catalyst, glacial acetic acid, only an error in the dosing rate was considered as it was used directly from the bottle and there was no sample preparation procedure involved. For each experiment, the pump was filled at room temperature with 17.48376 $mol \cdot L^{-1}$ glacial acetic acid (Carlo Erba Reagents for analysis, certified 100.0%). To initiate the reaction, 4.91 mL of the acid were dosed into the reactor content within 36 seconds. During the dosing period, a maximum temperature change of 0.5°C was observed.

Any potential volume expansion or contraction due to the mixing of the reactant solutions was tested by the addition of appropriate volumes of all three compounds into a graduated cylinder. No deviation from volume additivity (15 + 15 + 5 = 35 mL) was detected during the course of the reaction within the scale of the cylinder (0.25 mL) or the overall uncertainty due to pipetting (± 0.20 mL).

3.4.2.2 Instrumentation

Experiments were carried out in the Combined Reaction Calorimeter (CRC.v4) [35], a small-scale reaction calorimeter that combines the principle of power compensation and heat balance. It allows a maximum volume of 50 mL. Dosing was done by a Jasco HPLC pump (model PU-1580). Power compensation of the CRC.v4 is achieved by means of a compensation heater made of Hastelloy immersed into the reaction solution. The sensitivity of this compensation heater allows this calorimeter to maintain highly isothermal conditions (± 0.04°C), a prerequisite in order to minimise temperature effects on the error propagation. The jacket temperature is kept constant by Peltier elements. For a detailed description of the reactor, refer to [35].

Mid-IR signals were monitored between 1200 and 1650 cm^{-1} at 4 cm^{-1} resolution by Mettler Toledo's ReactIR 4000 system comprised of an FT-IR spectrometer connected via a K4 conduit to an ATR-IR crystal directly built into the bottom of the reactor vessel. UV-vis signals were followed between 240 and 400 nm at 1 nm resolution by a Cary 50 UV-vis spectrometer (Varian) coupled to a Hellma ATR-UV dip probe (model 661.804) immersed into the reactor. UV-vis and mid-IR signals were recorded simultaneously every minute for a total of 150 minutes.

Figure 3.5 shows a representative example of absorbance spectra acquired in the UV-vis and mid-IR range (solvent THF as reference background). No measurements were taken during the dosing period of 36 seconds.

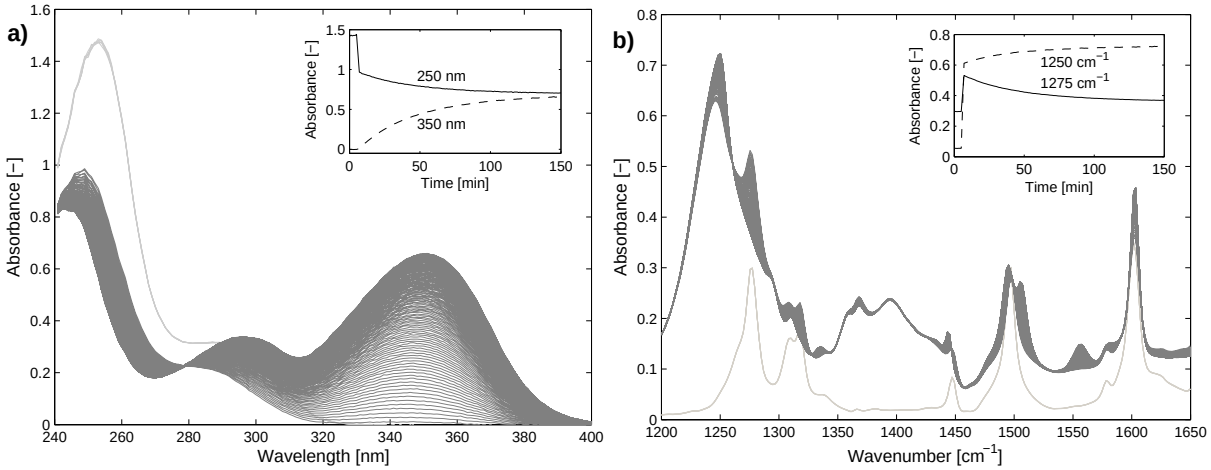

Figure 3.5: Time resolved UV-vis (a) and mid-IR (b) absorbance spectra before dosing (light gray curves) and after dosing (dark gray curves) with corresponding kinetic traces as insets. The sudden changes in absorbance in the mid-IR region are due to dilution and absorption from dosing of the catalyst acetic acid. In the UV-vis region only dilution effects are observed during dosing.

3.5 Results and discussion

In Section 3.3, equations for error propagation were derived in a general way, i.e. covering multi step mechanisms and multiple dosing events. This required the use of a vector **k** for the rate constants and a vector $\mathbf{f_s}$ for the flow rates. As examples discussed in this section only comprise one step reactions and one dosing event, vectors **k** and $\mathbf{f_s}$ collapse to scalars. However, for the sake of simplicity, the vector notation (boldface lowercase) is maintained. Thus, these vectors and their corresponding variances contain one element only.

3.5.1 Simulated data

3.5.1.1 Second order model ($A + B \rightarrow P$ or S)

Second order reactions are amongst the most common reactions (although often simplified to pseudo-first order for mathematical ease). In the following, we investigate in detail the individual contributions of Equations 3.9 or 3.10 that lead to a propagated error in a second order rate constant. Batch conditions were considered (no dosing), so there is no uncertainty to be propagated that corresponds to a dosing rate and the third term of Equation 3.10 can be disregarded. Thus, the variance of the second order rate constant $\boldsymbol{\sigma}_{\mathbf{k}}^2$ is the sum of the variance due to the residuals $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2$ and the variance due to the initial concentrations $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$.

To study the impact of these individual contributions to the variance of the fitted rate constant, the ratio of the initial concentrations of reactants *A* and *B* was varied between 0.01 and 100 while keeping the sum of the initial concentrations of the two reactants, $c_{tot} = c_{0,A} + c_{0,B}$, constant at 1 mol·L^{-1}. Data with strongly overlapped simulated component spectra were used here (product *P*, see Table 3.1 and Figure 3.3). As can be seen in Figure 3.6, the variance $\boldsymbol{\sigma}_{\mathbf{k}}^2$ is dominated by $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$, and $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2$ only plays a minor role at the chosen noise level of the simulated data ($\sigma_{\mathbf{y}}$ = 10^{-4}). Both $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$ as well as $\boldsymbol{\sigma}_{\mathbf{k}}^2$ show a sharp minimum at stoichiometric conditions ($c_{0,A}/c_{0,B}$ = 1). Slight deviations from stoichiometric conditions lead to an increase in $\boldsymbol{\sigma}_{\mathbf{k}}^2$ (and $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$) whereas larger deviations (i.e. when moving towards pseudo-first order conditions) to a subsequent decrease.

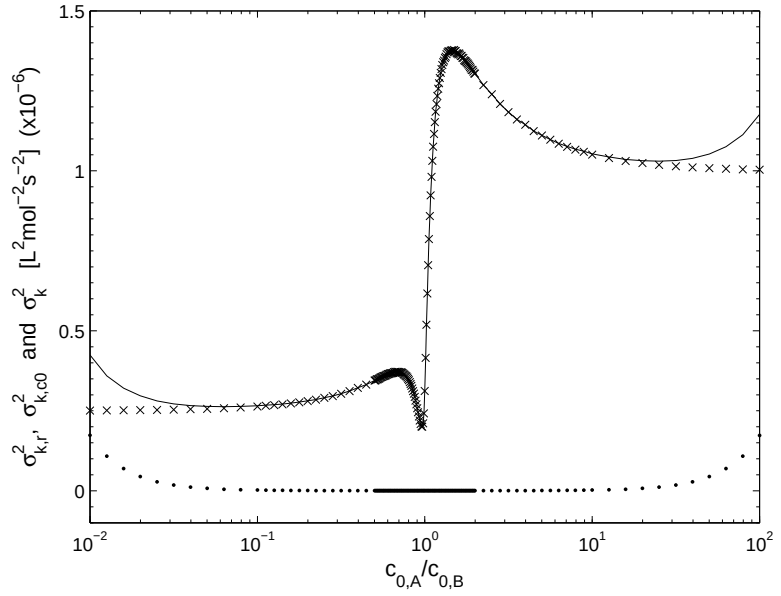

Figure 3.6: Individual contributions $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2$ (dotted line) and $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$ (crossed line) to the variance $\boldsymbol{\sigma}_{\mathbf{k}}^2$ (full line) of the fitted rate constant **k** as calculated from Equation 3.10 for the second order model $A+B \rightarrow P$ at various ratios $c_{0,A}/c_{0,B}$ of the initial concentrations ($c_{0,A} + c_{0,B}$ = 1 mol·L^{-1}) and imposed uncertainties of $\sigma_{c_{0,A}}$ = 0.2%·$c_{0,A}$ and $\sigma_{c_{0,B}}$ = 0.1%·$c_{0,B}$.

However, in very large excess of either species (strong pseudo-first order conditions), $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{r}}^2$ becomes more and more prominent and significantly adds to the total variance $\boldsymbol{\sigma}_{\mathbf{k}}^2$. This is due to the fact that the concentration of the limiting species is gradually decreasing as is the change in the absorbance. Eventually, it will reach the detection limit and no kinetics can be observed within the noise level of the instrument ($\sigma_{\mathbf{r}} \approx \sigma_{\mathbf{y}}$).

Note that a significant increase in the noise level $\sigma_{\mathbf{y}}$ of the simulated data matrix **Y** increases the contribution of $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{r}}$ on $\boldsymbol{\sigma}^2_{\mathbf{k}}$ and distorts the characteristic shape of $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ shown in Figure 3.6, no longer allowing for an interpretation. However, within one or two orders of magnitude in the noise level, $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ is hardly affected.

For further interpretation of the variance due to the initial concentrations (see Figure 3.6), the factors contributing to $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ are investigated separately. According to Equation 3.10, the variances in the initial concentrations $\boldsymbol{\sigma}^2_{\mathbf{c_0}}$ estimated by the experimenter weight the partial derivatives of **k** with respect to these initial concentrations. For this particular example, the variance $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ is $\left(\partial\mathbf{k}/\partial c_{0,A}\right)^2 \cdot \sigma^2_{c_{0,A}} + \left(\partial\mathbf{k}/\partial c_{0,B}\right)^2 \cdot \sigma^2_{c_{0,B}}$. Figure 3.7a shows that the vector $\boldsymbol{\sigma}^2_{\mathbf{c_0}} = [\sigma^2_{c_{0,A}}, \sigma^2_{c_{0,B}}]$ is responsible for the asymmetry of $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ in Figure 3.6, reflecting the difference in the relative errors imposed on the initial concentrations. Minima and maxima in $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ are due to the derivatives (see Figure 3.7b) and their interpretation or prediction is difficult as they are the result of two preceding optimisations (NGL/M) required for the numerical differentiation (see Figure 3.2).

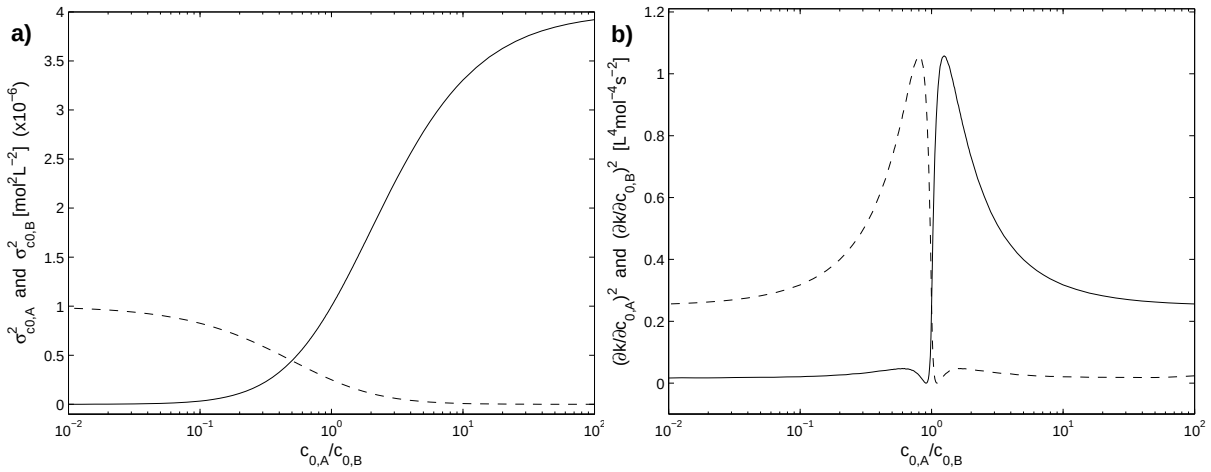

Figure 3.7: (a) Squared estimated uncertainties $\boldsymbol{\sigma}^2_{\mathbf{c_0}}$ (full line $\boldsymbol{\sigma}^2_{\mathbf{c_{0,A}}}$, dashed line $\boldsymbol{\sigma}^2_{\mathbf{c_{0,B}}}$) in the initial concentrations and (b) squared derivatives of the fitted rate constant **k** with respect to the initial concentrations (full line $\left(\partial\mathbf{k}/\partial c_{0,A}\right)^2$, dashed line $\left(\partial\mathbf{k}/\partial c_{0,B}\right)^2$) for the second order model $A + B \rightarrow P$ at various ratios $c_{0,A}/c_{0,B}$ and for $c_{0,A} + c_{0,B} = 1$ mol·L^{-1}.

A dramatic decrease of the spectral overlap (product *P* replaced by *S*, see Figure 3.3) does not result in any observable effect on the squared derivatives of Figure 3.7b and thus was not further investigated. However, a limited investigation on the effect of rate constant, conversion and sum of the initial concentrations on $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ revealed an impact on the squared derivatives $(\partial\mathbf{k}/\partial c_{0,A})^2$ and $(\partial\mathbf{k}/\partial c_{0,B})^2$. From an alternating variation of these parameters, the three following observations were made: (a) an increase of the rate constant narrows the minima and maxima of $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{c_0}}$ in Figure 3.6 and increases the derivatives; (b) with an increasing sum of the initial concentrations of the two reactants the derivatives also increase but their shape remains invariant; (c) both shape and magnitude of the derivatives are influenced by the conversion. The greater the conversion, the narrower becomes the minimum and the greater are the derivatives.

Non-linear optimisation problems generally do not have explicit solutions. Therefore it is impossible to predict $\partial\mathbf{k}/\partial\mathbf{c_0}$ without the numerical solution of the NGL/M algorithm. However, second order rate laws can be integrated explicitly and may be used to estimate some boundaries for these derivatives. In order to test this, the concentration profiles resulting from the explicit integration [36] of the corresponding differential equations (see Table 3.1) were rearranged for **k** and then derived analytically with respect to the initial concentrations (see Section 3.7.2 – Appendix). These obtained predicted derivatives do not reflect the impact of fitting the absorbance data, i.e. they do not depend on **Y**, and are only a function of the rate law and the corresponding set of ordinary differential equations. Surprisingly, the analytical squared derivatives of **k** with respect to the species in excess approximate the ones from the fitting towards strong pseudo-first order conditions (far right or far left of Figure 3.7b). In such conditions, the squared derivatives of **k** with respect to the limiting species and to the species in excess tend to zero and to $\left(\mathbf{k}/c_{\text{tot}}\right)^2$ respectively. These limits coincide with the analytical solution presented in the Section 3.7.2 (Appendix). This is also the case at the stoichiometric ratio ($c_{0,A}/c_{0,B} = 1$) where the squared derivatives from the fitting can again be predicted by their analytical solution, i.e. $\left(\mathbf{k}/c_{\text{tot}}\right)^2$ for *A* and *B* (see Figure 3.7b).

Despite these limits, it is not possible to approximate $\partial\mathbf{k}/\partial\mathbf{c_0}$ by an analytical equation due to an unpredictable impact of the non-linear fitting. In particular, the analytical derivation from the rate law requires breaking the continuity of the derivatives towards stoichiometric conditions, i.e. there are two different explicit functions that define the squared derivatives at stoichiometric ($c_{0,A} = c_{0,B}$) and non-stoichiometric ($c_{0,A} \neq c_{0,B}$) conditions. As the analytical squared derivatives with respect to the species in excess tend towards infinity approaching stoichiometric conditions, the two maxima close to the stoichiometric point of Figure 3.7b can be attributed to the fitting.

Figure 3.6 and Figure 3.7b suggest optimum experimental conditions when the two reactants are mixed in stoichiometric ratio. At this ratio, the impact of the errors in the initial concentrations on **k** is at its minimum. However, due to the fairly steep increase in the vicinity of the stoichiometric ratio (depending on the magnitude of the rate constant, the total concentration and conversion, as discussed above), it is preferable to choose pseudo-first order conditions to perform the experiments. Each squared derivative of Figure 3.7b is weighted by its associated variance given in Figure 3.7a to produce the overall variance profile for the fitted rate constant of Figure 3.6. So, the species with the lowest associated uncertainty in its initial concentration should be used in excess (here species *B*). Naturally, these findings go inline with experimental conditions often intuitively followed by kineticists.

3.5.1.2 Third order model (*A* + *B* + *C* → *P* + *C*)

Due to collision theory, elementary reactions of third order are rather unlikely. Nonetheless, empirical mechanisms of this order can be observed and explained by a steady state approximation of a more complex reaction [37]. A typical example taken from homogenous catalysis is a fast pre-equilibrium between a reactant *A* and a catalyst *C* to form a steady state complex *AC* that subsequently reacts with another reactant *B* to form a product *P* and to regenerate the catalyst *C*.

$$\begin{aligned} A + C &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AC \\ AC + B &\xrightarrow{k_2} P + C \end{aligned} \tag{3.13}$$

Application of the steady state approximation on *AC* under the condition that $k_{-1} \gg k_2 \cdot c_{t,B}$ results in an observed third order reaction of the form:

$$A + B + C \xrightarrow{k} P + C \quad \text{with} \quad \frac{\mathrm{d}c_{t,P}}{\mathrm{dt}} = k \cdot c_{\mathrm{t},A} c_{\mathrm{t},B} c_{\mathrm{t},C} \tag{3.14}$$

where $k = k_1 k_2 / k_{-1}$ is the observed third order rate constant. Under semi-batch conditions, e.g. when catalyst *C* is dosed, Equation 3.14 requires a minor adaptation to take into account dosing and dilution (see Table 3.1).

The proposed method of error propagation could be successfully applied and validated for semi-batch data simulated by this third order mechanism and subsequent non-linear optimisations of the rate constant, starting from various initial concentrations and flow rates (for the dosed species *C*) normally distributed around their true simulated values (Monte-Carlo sampling). The procedure for the Monte-Carlo sampling has been described at the end of section 3.4.1. The fitted rate constant and its associated standard deviation ($1.75(2) \pm 0.02(2) \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$) predicted by error propagation (Equation 3.10) is in perfect agreement with the mean and standard deviation ($1.75(2) \pm 0.02(3) \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$) of the 10^4 fitted rate constants resulting from the Monte-Carlo sampling. This demonstrates the formal correctness and the accuracy of the developed algorithm to propagate the errors on the initial concentrations and on the flow rate.

It is interesting to note that $\boldsymbol{\sigma}^2_{\mathbf{k},\mathbf{f_s}}$, the variance due to the dosing rate, contributes the most to the total predicted variance $\boldsymbol{\sigma}^2_{\mathbf{k}}$ and represents the main source of uncertainty ($\approx$ 94%) whereas the variance due to the initial concentrations only accounts for $\approx$ 6%. The contribution of the variance due to the residuals (0.004%) is basically negligible.

3.5.2 Experimental data

This section compares the uncertainties in the fitted rate constant based on one experiment only predicted by Equation 3.10 with the real one obtained from a collection of 17 independent experiments.

The acid catalysed reaction of phenylhydrazine with benzophenone (Figure 3.4) was repeatedly studied for this purpose by UV-vis and mid-IR spectroscopy. For batch conditions under acidic excess, it has previously been shown that this reaction follows an overall second order rate law, i.e. first order in both reactants [31]. When the catalyst is dosed (semi-batch conditions), the third order rate law as described in the previous section and in Table 3.1 (*A*: phenylhydrazine, *B*: benzophenone, *C*: acetic acid, *P*: benzophenone phenylhydrazone) becomes applicable. Note that water, as the by-product, has not been included into the kinetic analysis as it is not required for the rate law and as its spectrum cannot be separated from the pure spectrum of product *P* due to parallel formation.

Figure 3.8 compares fitted and measured kinetic absorbance traces at selected wavelengths and wavenumbers for a typical experiment. The sudden change in absorbance after 6 minutes corresponds to the dosing of acetic acid. Fits were generally good with typical standard deviations in the residuals, $\sigma_r \approx 5 \cdot 10^{-3}$ (UV-vis) and $5 \cdot 10^{-4}$ (mid-IR) explaining more than 99.9% of the total variance for both signals.

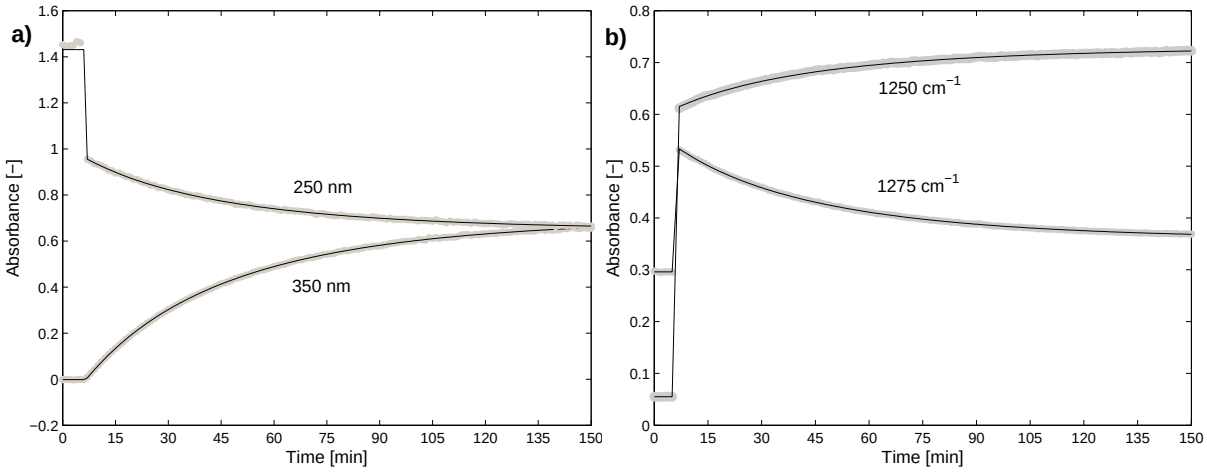

Figure 3.8: Comparison of fitted (lines) and measured (circles) absorbance traces in UV-vis (a) and in mid-IR (b). Spectra between 0 and 6 minutes are recorded before the dosing of the catalyst.

Fitted pure component spectra are presented in Figure 3.9 for a typical experiment. Note that these spectra are normalised to unity concentrations but not to unity path length as this distance is difficult to estimate for ATR probes.

Because of the intrinsic rank deficiency in the concentration profiles for third order rate laws, only linear combinations of the pure component spectra of *A*, *B* and *P* can be obtained [38,39]. Note that the dosed catalyst *C* (acetic acid), although being a low UV-vis absorber ($\lambda_{max} \approx 207$ nm), was also included as an absorbing component in order to allow for unavoidable small baseline shifts during the dosing event (which also affects the mid-IR spectrum of acetic acid). Due to rank deficiency mentioned above, only the pure spectra of acetic acid in UV-vis and mid-IR are resolved and can be compared to independently measured spectra for validation. They were in very good agreement for both UV-vis and mid-IR.

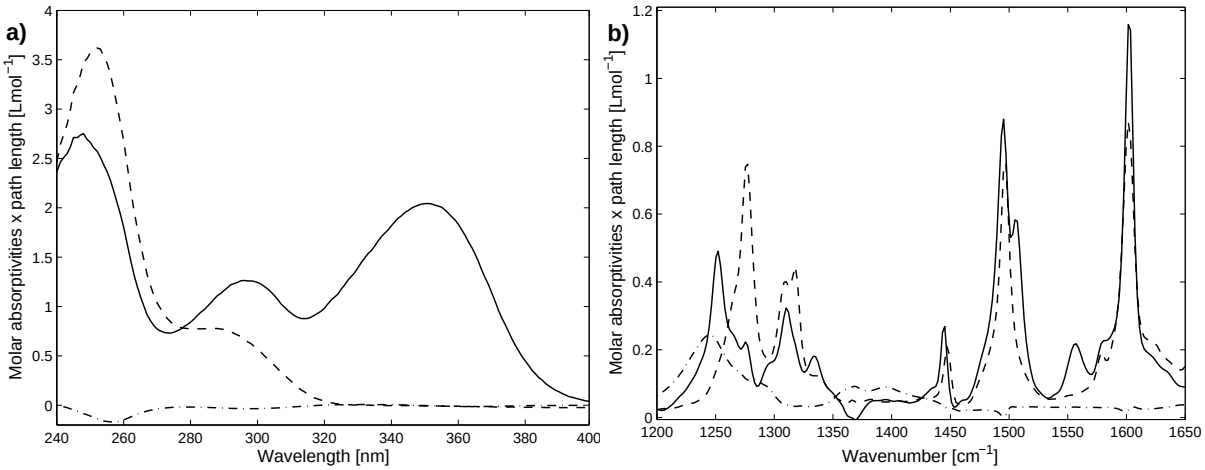

Figure 3.9: Fitted pure component spectra for benzophenone phenylhydrazone (solid line), for benzophenone (dashed line), both combined with the spectrum of phenylhydrazine, and resolved pure spectrum of acetic acid (dashed dot line) in UV-vis (a) and mid-IR (b).

Table 3.2 summarises all fitted rate constants for the 17 kinetic UV-vis and mid-IR experiments and their associated uncertainties with (Equation 3.10) and without (Equation 3.8) propagating the errors in the initial concentrations and in the flow rate.

An analysis of significance was performed using a z-test on the calculated rate constants. It showed that some few calculated rate constants were just close to the significance level of 95% confidence. Removing these values from such a small statistical sample (17 experiments) is arguable and does not affect the subsequent conclusions that can be made on the entire sample. This is why the entire statistical sample was kept intact.

Considering the 95% confidence limit, the fitted rate constants and all associated uncertainties in Table 3.2 are reproducible for both UV-vis and mid-IR. As expected, the uncertainties of the individual rate constants predicted by error propagation are significantly larger than those without error propagation.

Table 3.2: Fitted third order rate constants and associated uncertainties calculated with and without error propagation for the reaction of benzophenone with phenylhydrazine under semi-batch conditions studied by UV-vis and mid-IR spectroscopy. Numbers in brackets indicate the first insignificant digit.

	UV-vis			mid-IR		
	k [a)]	σ_k without error propagation [a), b)]	σ_k with error propagation [a), c)]	**k** [a)]	σ_k without error propagation [a), b)]	σ_k with error propagation [a), c)]
# 1	1.75(6)	0.003	0.02(2)	1.74(5)	0.002	0.02(2)
# 2	1.76(6)	0.003	0.02(3)	1.73(0)	0.001	0.02(2)
# 3	1.74(4)	0.003	0.02(2)	1.66(1)	0.003	0.02(2)
# 4	1.75(7)	0.003	0.02(2)	1.72(3)	0.004	0.02(2)
# 5	1.75(2)	0.003	0.02(2)	1.74(8)	0.002	0.02(2)
# 6	1.76(1)	0.003	0.02(2)	1.70(6)	0.003	0.02(1)
# 7	1.78(8)	0.004	0.02(4)	1.76(8)	0.002	0.02(2)
# 8	1.85(2)	0.004	0.02(5)	1.83(5)	0.002	0.02(3)
# 9	1.76(2)	0.004	0.02(3)	1.78(1)	0.004	0.02(2)
# 10	1.76(7)	0.004	0.02(3)	1.78(4)	0.002	0.02(3)
# 11	1.79(2)	0.004	0.02(4)	1.81(2)	0.002	0.02(3)
# 12	1.78(6)	0.004	0.02(4)	1.73(7)	0.004	0.02(2)
# 13	1.76(1)	0.004	0.02(3)	1.75(5)	0.002	0.02(2)
# 14	1.77(7)	0.004	0.02(3)	1.77(4)	0.002	0.02(2)
# 15	1.75(5)	0.004	0.02(3)	1.71(3)	0.003	0.02(2)
# 16	1.71(0)	0.004	0.02(3)	1.62(4)	0.002	0.02(0)
# 17	1.77(8)	0.004	0.02(3)	1.67(2)	0.003	0.02(1)

a) in $L^2 \cdot mol^{-2} \cdot s^{-1} \times 10^{-4}$

b) based on Equation 3.8.

c) based on Equation 3.10.

As shown in Table 3.3, the mean rate constants for UV-vis ($1.76(8) \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$) and for mid-IR ($1.73(9) \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$) are the same within their experimental standard deviation, and are comparable to previously published values reported for the same temperature under batch conditions [31]. Differences could be due to discrepancies in the real nominal temperature, fluctuations thereof, different experimental design (batch versus semi-batch conditions) and a different method for kinetic data analysis (PLS-calibration with subsequent direct fitting of concentration profiles).

Table 3.3: Means of all individually fitted rate constants (columns 2 and 5 of Table 3.2) compared to their literature values obtained from UV-vis and mid-IR spectroscopy. Standard deviations associated with columns 2 and 5 of Table 3.2 are compared to the mean uncertainties predicted by error propagation (columns 4 and 7 of Table 3.2). Numbers in brackets indicate the first insignificant digit.

	UV-vis		mid-IR	
	$\mathbf{k}$ [a]	$\boldsymbol{\sigma}_\mathbf{k}$ [a]	$\mathbf{k}$ [a]	$\boldsymbol{\sigma}_\mathbf{k}$ [a]
Experimental	1.76(8)	0.02(8)	1.73(9)	0.05(4)
Predicted by error propagation	-	0.02(3)	-	0.02(2)
Literature	1.40 [b]	[c]	1.51 [b]	[c]

[a] in $L^2 \cdot mol^{-2} \cdot s^{-1} \times 10^{-4}$

[b] for comparison, pseudo-second order rate constants of reference [31] have been recalculated to third order rate constants by division with the excess catalyst concentration given in [31]

[c] not available

For UV-vis spectroscopy, the uncertainty predicted by error propagation ($\boldsymbol{\sigma}_\mathbf{k}$ = 0.02(3)·10^{-4} L^2·mol^{-2}·s^{-1}) covers more than 80% of the experimental standard deviation over all individually fitted rate constants ($\boldsymbol{\sigma}_\mathbf{k}$ = 0.02(8)·10^{-4} L^2·mol^{-2}·s^{-1}). This suggests that the approach used to propagate uncertainties in initial concentrations and flow rate is suitable. Possible sources for the remainder could be attributed to effects such as minor instrumental drifts (e.g. in baseline or flow rate) that have not been modelled. Additionally, minor fluctuations in temperature ($\sigma_\mathbf{T}^2 \approx 0.04$°C) have an impact. Assuming, for example, an activation energy between 50 and 100 kJ·mol^{-1} (to our knowledge, the true value is not known), a corresponding contribution incorporated in Equation 3.10 due to fluctuations in temperature $\mathbf{T}$, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{T}}^2 = \mathrm{diag}\left(\left(\partial\mathbf{k}/\partial\mathbf{T}\right)^\mathrm{T} \sigma_\mathbf{T}^2 \left(\partial\mathbf{k}/\partial\mathbf{T}\right)\right)$ with $\partial\mathbf{k}/\partial\mathbf{T}$ determined from Arrhenius' law, would add between 2% and 8% to the total predicted error $\boldsymbol{\sigma}_\mathbf{k}$. Note that the lack of information in manufacturer's uncertainties in the purity of chemicals did not allow them to be taken into account in the error estimate of the initial concentrations from the sample preparation.

For mid-IR measurements, error propagation ($\boldsymbol{\sigma}_\mathbf{k}$ = 0.02(2)·10^{-4} L^2·mol^{-2}·s^{-1}) only explains approximately 40% of the experimental uncertainty ($\boldsymbol{\sigma}_\mathbf{k}$ = 0.05(4)·10^{-4} L^2·mol^{-2}·s^{-1}) in the fitted rate constant. All sources explaining the remaining contributions to the experimental uncertainty listed for UV-vis certainly also apply to mid-IR spectroscopy.

One possible reason to explain the lower error predictability observed for mid-IR (40%) compared to UV-vis (80%) could be a generally larger sensitivity to temperature for mid-IR spectroscopy. Another explanation could be a larger deviation from the silent assumption of normally distributed and homoscedastic (i.e. constant standard deviation) noise in the experimental mid-IR absorbance data used for the least squares fitting [40]. This last statement was supported by multiple kinetic fittings at single wavelengths/wavenumbers (with similar absorbance changes) for both UV-vis and mid-IR data leading to a lower wavelength dependency of the residual noise level and of the fitted rate constant for UV-vis.

The residual standard deviation of the noise in mid-IR spectroscopy is possibly less constant along the wavenumbers. This could be counteracted by an appropriate weighting of the residuals in Equation 3.3 according to a time and wavenumber dependent standard deviation in **Y**. This method is often called χ^2 (chi square) fitting. The χ^2 fitting will result in another experimental standard deviation in the rate constants (over all 17 experiments) that needs to be compared with their predicted uncertainties for each individual measurement that will also be affected. Thus, the percentage of experimental standard deviation explained by error propagation will change accordingly. A similar effect has already been discussed by [19]. In practice, however, for time resolved multivariate absorbance measurements it can be fairly difficult to reliably determine the individual standard deviations in **Y** as the access to the instrument control is generally rather limited by the instrument manufacturer.

The total predicted variance $\boldsymbol{\sigma}_{\mathbf{k}}^2$ of the rate constant fitted to the UV-vis or mid-IR measurements can be broken down into its individual contributions ($\boldsymbol{\sigma}_{\mathbf{k,r}}^2$, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{f_s}}^2$) given in Equation 3.10. For both UV-vis and mid-IR, the variance due to the residuals, $\boldsymbol{\sigma}_{\mathbf{k,r}}^2$, has the lowest contribution (≈ 2%), followed by the variance due to the initial concentrations, $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{c_0}}^2$ that explains approximately 6% of the total variance. Clearly the major contribution (≈ 92%) is attributed to $\boldsymbol{\sigma}_{\mathbf{k},\mathbf{f_s}}^2$, the variance due to the flow rate. This is in good agreement with the results obtained for simulated data based on the same model discussed in Section 3.5.1.2. Particularly for mid-IR, the real individual contributions might be somewhat different due to possible heteroscedastic noise as outlined in the previous paragraph.

3.6 Conclusion

The impact of uncertainties in the initial concentrations and dosing rate (batch and/or semi-batch) on the error of rate constants fitted by multivariate non-linear regression of spectroscopic data has been studied on simulated and experimental data following second and formal third order rate laws. For this, a rigorous approach based on classical error propagation was developed and theoretically validated with simulated data by a 'brute force' Monte-Carlo sampling. Rigorous error propagation required significantly less computation time compared to the Monte Carlo procedure (typically seconds compared to hours on a modern personal computer).

Based on simulated data for a second order rate law under batch conditions, the effect of the uncertainties in the initial reactant concentrations on the predicted error of the fitted rate constant was thoroughly investigated. It was shown that the predicted error in the fitted second order rate constant is minimal when the reactants are mixed in exact stoichiometric amounts, or if the species with the lowest associated uncertainty in its initial concentration is in excess (pseudo-first order conditions). These findings have an immediate application in the optimum experimental design of second order reactions.

The reaction of benzophenone with phenylhydrazine was repeatedly investigated by UV-vis and mid-IR spectroscopy under semi-batch conditions dosing the catalyst (acetic acid). Each measurement was analysed individually using a third order rate law and the standard deviation of each fitted rate constant was predicted by the proposed method of error propagation. For UV-vis, each individual predicted standard deviation covered more than 80% of the experimental standard deviation over all individually fitted rate constants. For mid-IR, the prediction was only capable to explain approximately 40% of the experimental uncertainty in the fitted rate constant. Possible reasons for this lower error predictability could be a larger sensitivity to temperature and/or stronger deviations from an assumed normally distributed and homoscedastic noise for mid-IR compared to UV-vis.

Amongst the individual contributions towards the predicted variance in the rate constant, the dosing rate has by far the largest impact (≈ 92%), followed by the initial concentrations (≈ 6%) and the variance due to the residuals (≈ 2%). These individual contributions are in good agreement with the results obtained for data simulated and analysed under similar conditions.

The possibility to reasonably predict the error in the rate constant based on one single multivariate kinetic measurement was demonstrated. The proposed method of error propagation is simple to implement and fast to perform in order to receive a quick estimate of the error in the rate constant and the individual contributions to this error. In practice, however, errors obtained from error propagation should always be compared with the experimental standard deviation obtained from a reasonable number of replicates. One obvious advantage of the rigorous error propagation is its ability to pinpoint the major source of error and to quantify their impact onto the fitted rate constant. In the presented case study, this would clearly be the dosing rate of the pump. Furthermore, Equation 3.10 can straightforwardly be extended to also cover other sources of uncertainty provided reasonable error estimates are available and can be propagated; for example, manufacturer's uncertainties in the purity of the chemicals, if available, could be included in Equation 3.10 as a fourth term in the error propagation.

3.7 Appendix

3.7.1 Error estimation in the initial concentrations due to sample preparation

Errors in the sample preparation (as described in Section 3.4.2.1) for the reactants phenylhydrazine (*A*) and benzophenone (*B*) were determined by propagating the following uncertainties of the equipment:

1. Weighing: $\sigma_{m_A} = \sigma_{m_B} = 0.0001$ g
2. Filling the volumetric 50 mL flasks: $\sigma_{v_A^{\text{flask}}} = \sigma_{v_B^{\text{flask}}} = 0.06$ mL
3. Pipetting (pip1: 10 mL, pip2: 5mL): $\sigma_{v_A^{\text{pip1}}} = \sigma_{v_B^{\text{pip1}}} = \sigma_{v_A^{\text{pip2}}} = \sigma_{v_B^{\text{pip2}}} = 0.04$ mL

For description of the variables we refer to the list of symbols. The initial concentrations ($c_{0,A}$ and $c_{0,B}$) in the reactor were calculated according to:

$$c_{0,A} = \frac{c_{0,A}^{\text{flask}}\left(v_A^{\text{pip1}} + v_A^{\text{pip2}}\right)}{v_{\text{tot}}} \quad \text{and} \quad c_{0,B} = \frac{c_{0,B}^{\text{flask}}\left(v_B^{\text{pip1}} + v_B^{\text{pip2}}\right)}{v_{\text{tot}}} \tag{3.15}$$

with $c_{0,A}^{\text{flask}} = m_A \Big/ \left(M_A\ v_A^{\text{flask}}\right)$, $c_{0,B}^{\text{flask}} = m_B \Big/ \left(M_B\ v_B^{\text{flask}}\right)$ and $v_{\text{tot}} = v_A^{\text{pip1}} + v_A^{\text{pip2}} + v_B^{\text{pip1}} + v_B^{\text{pip2}}$

Rigorous error propagation is performed via the total derivatives of $c_{0,A}$ and $c_{0,B}$:

$$\sigma_{c_{0,A}}^2 = \left(\frac{\partial c_{0,A}}{\partial m_A}\right)^2 \sigma_{m_A}^2 + \left(\frac{\partial c_{0,A}}{\partial v_A^{\text{flask}}}\right)^2 \sigma_{v_A^{\text{flask}}}^2 + \left(\frac{\partial c_{0,A}}{\partial v_A^{\text{pip1}}}\right)^2 \sigma_{v_A^{\text{pip1}}}^2 + \left(\frac{\partial c_{0,A}}{\partial v_A^{\text{pip2}}}\right)^2 \sigma_{v_A^{\text{pip2}}}^2 + \left(\frac{\partial c_{0,A}}{\partial v_B^{\text{pip1}}}\right)^2 \sigma_{v_B^{\text{pip1}}}^2 + \left(\frac{\partial c_{0,A}}{\partial v_B^{\text{pip2}}}\right)^2 \sigma_{v_B^{\text{pip2}}}^2 \tag{3.16}$$

$$\sigma_{c_{0,B}}^2 = \left(\frac{\partial c_{0,B}}{\partial m_B}\right)^2 \sigma_{m_B}^2 + \left(\frac{\partial c_{0,B}}{\partial v_B^{\text{flask}}}\right)^2 \sigma_{v_B^{\text{flask}}}^2 + \left(\frac{\partial c_{0,B}}{\partial v_B^{\text{pip1}}}\right)^2 \sigma_{v_B^{\text{pip1}}}^2 + \left(\frac{\partial c_{0,B}}{\partial v_B^{\text{pip2}}}\right)^2 \sigma_{v_B^{\text{pip2}}}^2 + \left(\frac{\partial c_{0,B}}{\partial v_A^{\text{pip1}}}\right)^2 \sigma_{v_A^{\text{pip1}}}^2 + \left(\frac{\partial c_{0,B}}{\partial v_A^{\text{pip2}}}\right)^2 \sigma_{v_A^{\text{pip2}}}^2 \tag{3.17}$$

The variances of *A* and *B* have therefore the following expressions:

$$\sigma_{c_{0,A}}^2 = \left[\left(\frac{\sigma_{m_A}}{m_A}\right)^2 + \left(\frac{\sigma_{V_A^{\text{flask}}}}{v_A^{\text{flask}}}\right)^2 + \left(\frac{1}{v_A^{\text{pip1}} + v_A^{\text{pip2}}} - \frac{1}{v_{\text{tot}}}\right)^2 \left(\sigma_{v_A^{\text{pip1}}}^2 + \sigma_{v_A^{\text{pip2}}}^2\right) + \left(\frac{1}{v_{\text{tot}}}\right)^2 \left(\sigma_{v_B^{\text{pip1}}}^2 + \sigma_{v_B^{\text{pip2}}}^2\right)\right] c_{0,A}^2 \tag{3.18}$$

$$\sigma_{c_{0,B}}^2 = \left[\left(\frac{\sigma_{m_B}}{m_B}\right)^2 + \left(\frac{\sigma_{V_B^{\text{flask}}}}{v_B^{\text{flask}}}\right)^2 + \left(\frac{1}{v_B^{\text{pip1}} + v_B^{\text{pip2}}} - \frac{1}{v_{\text{tot}}}\right)^2 \left(\sigma_{v_B^{\text{pip1}}}^2 + \sigma_{v_B^{\text{pip2}}}^2\right) + \left(\frac{1}{v_{\text{tot}}}\right)^2 \left(\sigma_{v_A^{\text{pip1}}}^2 + \sigma_{v_A^{\text{pip2}}}^2\right)\right] c_{0,B}^2 \tag{3.19}$$

3.7.2 Analytical derivatives of a second order rate constant with respect to the initial concentrations

The system of ODEs describing a second order rate law can be integrated explicitly in order to obtain the concentrations of *A* and *B* at any time t_X, i.e. the time required to reach conversion *X*. Subsequent rearrangement for the rate constant **k** leads to:

$$\mathbf{k} = \frac{1}{t_X\left(c_{0,B} - c_{0,A}\right)} \ln\left(\frac{c_{0,B} - c_{0,A}X}{c_{0,B}\left(1-X\right)}\right) \qquad \text{for } c_{0,A} < c_{0,B} \qquad (3.20)$$

$$\mathbf{k} = \frac{2X}{t_X\left(c_{0,A} + c_{0,B}\right)\left(1-X\right)} \qquad \text{for } c_{0,A} = c_{0,B} \qquad (3.21)$$

$$\mathbf{k} = \frac{1}{t_X\left(c_{0,B} - c_{0,A}\right)} \ln\left(\frac{c_{0,A}\left(1-X\right)}{c_{0,A} - c_{0,B}X}\right) \qquad \text{for } c_{0,A} > c_{0,B} \qquad (3.22)$$

The analytical partial derivatives of **k** with respect to $c_{0,A}$ and $c_{0,B}$ are given by the following expressions:

$$\left(\frac{\partial \mathbf{k}}{\partial c_{0,A}}\right) = \frac{\mathbf{k}}{\left(c_{0,B} - c_{0,A}\right)} + \frac{1}{t_X} \frac{X}{\left(c_{0,A}X - c_{0,B}\right)\left(c_{0,B} - c_{0,A}\right)}$$
$$\left(\frac{\partial \mathbf{k}}{\partial c_{0,B}}\right) = -\frac{\mathbf{k}}{\left(c_{0,B} - c_{0,A}\right)} - \frac{1}{t_X} \frac{c_{0,A}}{c_{0,B}} \frac{X}{\left(c_{0,A}X - c_{0,B}\right)\left(c_{0,B} - c_{0,A}\right)} \qquad \text{for } c_{0,A} < c_{0,B} \qquad (3.23)$$

$$\left(\frac{\partial \mathbf{k}}{\partial c_{0,A}}\right) = \left(\frac{\partial \mathbf{k}}{\partial c_{0,B}}\right) = -\frac{1}{t_X} \frac{2X}{\left(1-X\right)} \frac{1}{\left(c_{0,A} + c_{0,B}\right)^2} = -\frac{\mathbf{k}}{\left(c_{0,A} + c_{0,B}\right)} = -\frac{\mathbf{k}}{c_{\text{tot}}} \qquad \text{for } c_{0,A} = c_{0,B} \qquad (3.24)$$

$$\left(\frac{\partial \mathbf{k}}{\partial c_{0,A}}\right) = \frac{\mathbf{k}}{\left(c_{0,B} - c_{0,A}\right)} + \frac{1}{t_X} \frac{c_{0,B}}{c_{0,A}} \frac{X}{\left(c_{0,B}X - c_{0,A}\right)\left(c_{0,B} - c_{0,A}\right)}$$
$$\left(\frac{\partial \mathbf{k}}{\partial c_{0,B}}\right) = -\frac{\mathbf{k}}{\left(c_{0,B} - c_{0,A}\right)} - \frac{1}{t_X} \frac{X}{\left(c_{0,B}X - c_{0,A}\right)\left(c_{0,B} - c_{0,A}\right)} \qquad \text{for } c_{0,A} > c_{0,B} \qquad (3.25)$$

Towards strong pseudo-first order conditions in B ($c_{0,A} << c_{0,B}$):

$$\lim_{c_{0,A}\to 0}\left(\frac{\partial \mathbf{k}}{\partial c_{0,B}}\right)=\lim_{c_{0,A}\to 0}\ -\frac{\mathbf{k}}{\left(c_{0,B}-c_{0,A}\right)}-c_{0,A}\left(\frac{X}{t_X c_{0,B}\left(c_{0,A}X-c_{0,B}\right)\left(c_{0,B}-c_{0,A}\right)}\right)=-\frac{\mathbf{k}}{c_{0,B}}=-\frac{\mathbf{k}}{c_{\text{tot}}} \tag{3.26}$$

Towards strong pseudo-first order conditions in A ($c_{0,A} >> c_{0,B}$):

$$\lim_{c_{0,B}\to 0}\left(\frac{\partial \mathbf{k}}{\partial c_{0,A}}\right)=\lim_{c_{0,B}\to 0}\ \frac{\mathbf{k}}{\left(c_{0,B}-c_{0,A}\right)}+c_{0,B}\left(\frac{X}{t_X c_{0,A}\left(c_{0,B}X-c_{0,A}\right)\left(c_{0,B}-c_{0,A}\right)}\right)=-\frac{\mathbf{k}}{c_{0,A}}=-\frac{\mathbf{k}}{c_{\text{tot}}} \tag{3.27}$$

3.8 References

[1] B. Ma, P.J. Gemperline, E. Cash, M. Bosserman, E. Comas, J. Chemom., 17 (2003), 470-479.
[2] G. Puxty, U. Fischer, M. Jecklin, K. Hungerbühler, Chimia, 60 (2006), 605-610.
[3] A.E. Rubin, S. Tummala, D.A. Both, C.C. Wang, E.J. Delaney, Chemical Reviews, 106 (2006), 2794-2810.
[4] J. Workman, M. Koch, D. Veltkamp, Anal. Chem., 79 (2007), 4345-4363.
[5] E.R. Malinowski, Factor Analysis in Chemistry, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
[6] M. Maeder, Anal. Chem., 59 (1987), 527-530.
[7] J. Saurina, S. Hernandez-Cassou, R. Tauler, A. Izquierdo-Ridorsa, J. Chemom., 12 (1998), 183-203.
[8] J. Jaumot, R. Gargallo, R. Tauler, J. Chemom., 18 (2004), 327-340.
[9] A. de Juan, M. Maeder, M. Martinez, R. Tauler, Chemom. Intell. Lab. Syst., 54 (2000), 123-141.
[10] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst., 59 (2001), 19-31.
[11] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, Trac-Trends Anal. Chem., 26 (2007), 752-765.
[12] P. Geladi, B.R. Kowalski, Anal. Chim. Acta, 185 (1986), 1-17.
[13] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, Comput. Chem. Eng., 20 (1996), 1451-1465.
[14] M. Maeder, A.D. Zuberbühler, Anal. Chem., 62 (1990), 2220-2224.
[15] E. Furusjö, L.G. Danielsson, Anal. Chim. Acta, 373 (1998), 83-94.
[16] S. Bijlsma, D.J. Louwerse, A.K. Smilde, AIChE J., 44 (1998), 2713-2723.
[17] G. Puxty, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 81 (2006), 149-164.
[18] M. Maeder, Y.M. Neuhold, Chapter 7 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 218-256.
[19] M. Maeder, Y.M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, NL, 2007.
[20] P. Gemperline, G. Puxty, M. Maeder, D. Walker, F. Tarczynski, M. Bosserman, Anal. Chem., 76 (2004), 2575-2582.
[21] T. Kourti, Crit. Rev. Anal. Chem., 36 (2006), 257-278.
[22] K.Q. Levenberg, Appl. Math., 2 (1949), 164-168.
[23] D.W. Marquardt, J. Soc. Ind. Appl. Math., 11 (1963), 431-441.
[24] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C++ - The art of Scientific Computing, Second Edition, Cambridge University Press, New York, USA, 2005.
[25] P.R. Bevington, D.K. Robinson, Data reduction and error analysis for the physical sciences, Mcgraw-Hill, New York, USA, 2003.
[26] E. Furusjö, L.G. Danielsson, J. Chemom., 14 (2000), 483-499.
[27] V.M. Taavitsainen, H. Haario, J. Chemom., 15 (2001), 215-239.
[28] S. Bijlsma, H.F.M. Boelens, H.C.J. Hoefsloot, A.K. Smilde, J. Chemom., 16 (2002), 28-40.
[29] A.R. Carvalho, R.G. Brereton, T.J. Thurston, R.E.A. Escott, Chemom. Intell. Lab. Syst., 71 (2004), 47-60.

[30] V.M. Taavitsainen, H. Haario, M. Laine, J. Chemom., 17 (2003), 140-150.
[31] A.R. de Carvalho, M.D. Sanchez, J. Wattoom, R.G. Brereton, Talanta, 68 (2006), 1190-1200.
[32] R. Dyson, M. Maeder, G. Puxty, Y.M. Neuhold, Inorg. React. Mech., 5 (2003), 39-46.
[33] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007; http://www.mathworks.com
[34] J.F. Zhang, IEEE Trans. Rel., 55 (2006), 169-181.
[35] F. Visentin, S.I. Gianoli, A. Zogg, O.M. Kut, K. Hungerbühler, Org. Proc. Res. Dev., 8 (2004), 725-737.
[36] K.J. Laidler, J.H. Meiser, B.C. Sanctuary, Physical Chemistry, Houghton Mifflin, Boston, USA, 2003.
[37] J.H. Espenson, Chemical kinetics and reaction mechanisms, Mcgraw-Hill, New York, USA, 1995.
[38] R.N. Cochran, F.H. Horne, Anal. Chem., 49 (1977), 846-853.
[39] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemom. Intell. Lab. Syst., 33 (1996), 17-33.
[40] J. Tellinghuisen, Appl. Spectrosc., 54 (2000), 1208-1213.
[41] E. Sanchez, B.R. Kowalski, J. Chemom., 4 (1990), 29-45.
[42] R. Bro, Chemom. Intell. Lab. Syst., 38 (1997), 149-171.
[43] R. Tauler, A. de Juan, Chapter 11 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 417-474.
[44] M. Amrhein, Reaction and flow variants/invariants for the analysis of chemical reaction data, Doctoral Thesis 1861, EPF Lausanne, Switzerland, 1998.
[45] Y.M. Neuhold, M. Maeder, J. Chemom., 16 (2002), 218-227.
[46] J.J. Kankare, Anal. Chem., 42 (1970), 1322-&.
[47] M. Garland, E. Visser, P. Terwiesch, D.W.T. Rippin, Anal. Chim. Acta, 351 (1997), 337-358.
[48] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemom. Intell. Lab. Syst., 46 (1999), 249-264.
[49] A.E. Croce, J. Chem. Educ., 79 (2002), 506-509.
[50] H. Maskill, Educ. Chem., 21 (1984), 122-123.
[51] J. Billeter, Y.M. Neuhold, L.L. Simon, G. Puxty, K. Hungerbühler, Chemom. Intell. Lab. Syst., 93 (2008), 120-131.
[52] D.C. Lay, Linear Algebra and Its Applications, Third Edition, Addison Wesley Higher Education, Boston, USA, 2003.
[53] M. Maeder, Y.M. Neuhold, A. Olsen, G. Puxty, R. Dyson, A. Zilian, Anal. Chim. Acta, 464 (2002), 249-259.
[54] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta, 396 (1999), 231-240.
[55] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurenczy, M. Maeder, A.E. Merbach, A.D. Zuberbühler, Anal. Chim. Acta, 298 (1994), 193-201.
[56] R. Dyson, M. Maeder, Y.M. Neuhold, G. Puxty, Anal. Chim. Acta, 490 (2003), 99-108.
[57] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst., 81 (2006), 82-93.

CHAPTER 4

Systematic prediction of linear dependencies in the concentration profiles and implications on the kinetic hard-modelling of spectroscopic data

This chapter (except Section 4.6.5 – Appendix) was published as an article under the reference:

J. Billeter, Y.-M. Neuhold, K. Hungerbühler
Chemometrics and Intelligent Laboratory Systems, 95 (2009), 170-187
http://dx.doi.org/10.1016/j.chemolab.2008.10.002

* * *

4.1 Abstract

A novel method is presented for the systematic identification of the minimum requirements regarding mathematical pre-treatment, a priori information, or experimental design, in order to allow optimising rate constants and pure component spectra associated with a kinetic model via multivariate kinetic hard-modelling of spectroscopic data. Rank deficiencies in the kinetic concentration matrix represent a major problem for the calibration free method developed by Maeder and Zuberbühler, as its pseudo-inverse, required for the optimisation process, is not defined. In this contribution, the underlying linear dependencies in the concentration profiles are systematically elucidated and appropriate strategies are discussed in order to break them. Also, conditions are predicted for which full spectral resolution can be expected. The method is based on the kernel of a time invariant augmented matrix covering potential rank deficiency due to stoichiometry and rate laws, also relevant for the concentration matrix. Compared to employing the full concentration matrix, this augmented matrix does not require a numerical integration of the differential equations describing the kinetic model and thus can easily be set up. The kernel can be calculated numerically by Singular Value Decomposition (SVD) or determined in a symbolical way, the latter allowing the detection of particular stoichiometric conditions leading to spectral resolution of species. The capabilities of the method are demonstrated analysing three kinetic mechanisms of increasing complexity covering consecutive and parallel reactions.

4.2 Introduction

In recent years, chemometric methods dedicated to the analysis of evolutionary multivariate data (e.g. as in kinetic and equilibrium investigations) measured by spectroscopy (e.g. mid-IR, UV-vis, Raman or fluorescence) have considerably progressed in chemistry and chemical engineering [1-10]. Kinetic hard-modelling is one of these chemometric methods. It is used to determine the kinetic parameters (e.g. rate constants) of a chemical reaction [2,7,10-14]. As it is based on a hard model, kinetic hard-modelling has excellent predictive capabilities, as long as all chemical and physical effects are covered by the model [15]. For a good overview of the different hard-modelling techniques, we refer to [16,17].

Under the assumption of bi-linearity, time and wavelength resolved spectroscopic data can be decomposed according to Beer's law into the concentration profiles and the molar spectra of the pure components. In kinetic hard-modelling, these concentration profiles are usually obtained by numerical integration of the rate laws describing the underlying kinetic hard model.

Maeder and Zuberbühler [2] suggested that the associated kinetic parameters are adjusted by non-linear optimisation, so that the residuals between the measured absorbances and those modelled by Beer's law become minimal in the least-squares sense. In this calibration free method, sometimes referred to as kinetic fitting by implicit direct calibration [16,17], the pure component spectra, which are linear parameters, are eliminated from the non-linear optimisation and determined at each iteration by linear regression using the pseudo-inverse of the matrix comprising the concentration profiles. If the number of reactive species equals the rank of the concentration matrix, all concentration profiles are linearly independent and pure component spectra can be resolved. However, if the rank of the concentration matrix is lower than the number of species, i.e. if some concentration profiles are linearly dependent, the pseudo-inverse and thus the pure component spectra cannot be directly computed. In such case, there are at least five different strategies, as described in Table 4.1, to treat rank deficiency problems in the concentration matrix and to allow the fitting of pure component spectra [18].

Table 4.1. Five strategies to treat linear dependencies in the kinetic concentration profiles without affecting the non-linear optimisation, and their effects on the fitted pure component spectra.

Strategy		Effect on the fitted pure component spectra
1)	Model reduction by defining some absorbing species as uncoloured [a)] for data analysis [10]	Only partial spectral resolution. Linear combinations of the true pure component spectra are obtained.
2)	Model reduction by providing known pure component spectra for data analysis [10,19]	Complete spectral resolution if the appropriate pure component spectra are provided.
3)	Rank augmentation by dosing one or more species [20,21]	Complete spectral resolution if the appropriate species are dosed.
4)	Rank augmentation by varying initial concentrations and performing a 3-way or second-order global analysis [14,22-24]	Complete spectral resolution if appropriate experiments are performed.
5)	Perform a tri-linear experiment and treat the linear dependencies, for example, with Strategy (1) [18]	Complete spectral resolution due to an additional dimension of resolution, e.g. chromatography.

[a)] The term uncoloured is borrowed from UV-vis spectroscopy. In this context, uncoloured means that, although a species is a true absorber, it is treated as non-absorbing for mathematical reasons.

The rank of the spectral measurement matrix has been studied by Kankare [25], Cochran et al [20], Amrhein et al [21] and Garland et al [26] under batch and semi-batch conditions (Strategy 3) and for the concatenation of experiments measured under different initial conditions (Strategy 4). However, the effects of defining a specific absorbing species as uncoloured (Strategy 1), of providing a specific component spectrum (Strategy 2), of dosing a specific species (Strategy 3) or of varying the initial concentrations (Strategy 4) on the rank of the kinetic concentration matrix have not yet been studied in literature. These effects are of importance for the experimental design of kinetic experiments as they will reveal the required information to be provided or the additional experiments to be performed in order to allow a hard kinetic analysis according to the method of Maeder and Zuberbühler, and obtain partial or full spectral resolution. Until now, the appropriate species to be included in these strategies are usually selected by experience, intuition, or trial and error.
If the computation of the pseudo-inverse of the concentration matrix generates an error, it indicates a still rank deficient model and a wrong species selection within the corresponding strategy.

In this article, we propose a method for a systematic experimental and data analytical design of bi-linear spectroscopic kinetic measurements that allows identifying the species to be incorporated in strategies (1) to (4). In this contribution, we do not consider the kinetic analysis of tri-linear data – Strategy (5) – as this has been discussed in literature as a special case of Strategy (1), leading to full spectral resolution due to an additional dimension of resolution [18]. If Strategy (1) is used, the method optionally includes the calculation of the linear combinations of the true pure component spectra, i.e. the coefficients by which the true (resolved) pure component spectra need to be weighted in order to obtain a reduced set of fitted (unresolved) component spectra of the species treated as coloured (i.e. not uncoloured).

The proposed method is based on the analysis of the kernel of an augmented matrix that has the same kernel as the original concentration matrix. This augmented matrix can be deduced from a general mass balance, which is identical to the general equation proposed by Amrhein et al and published as the factorisation of concentration data [27]. In contrast to Amrhein et al that assume independent reactions, which may be difficult to identify for complex mechanisms, our systematic method is based on the entire set of kinetic reactions. The proposed method only requires the reduction of the entire set of kinetic reactions to its independent reactions if the linear combinations of the true pure component spectra are desired, as an option with Strategy (1). For this reduction, we also propose a systematic method.

The analysis of this augmented matrix avoids unnecessary and time-consuming numerical integrations to determine the whole concentration matrix and allows the experimental design of chemical reactions, even if the associated rate constants are not yet known, i.e. before they are optimised. The proposed method is presented using three kinetic examples of increasing complexity. From this, appropriate experimental designs (strategies) are suggested.

4.3 Theory

In kinetic modelling of spectroscopic data, Beer's law is used, under the assumption of bi-linearity, to decompose a measured absorbance signal into the concentrations and the molar spectra of the pure components. This law can elegantly be written in matrix notation.

$$\mathbf{Y} = \mathbf{CA} + \mathbf{R} \tag{4.1}$$

Y (*nt* x *nw*) denotes the measured time and wavelength resolved spectroscopic data, **C** (*nt* x *ns*) the concentration profiles and **A** (*ns* x *nw*) the pure component spectra. The matrix dimensions are defined using *nt* as the number of reaction times, *nw* as the number of wavelengths and *ns* as the total number of species. The deviations from the product of **C** and **A** defined by Beer's law are captured in a matrix **R** (*nt* x *nw*) of residuals.

Kinetic hard-modelling involves a chemical hard model to describe the concentration profiles of the reactive species. This chemical model can be uniquely defined by matrices **E** (*nr* x *ns*) and **P** (*nr* x *ns*), both comprised of positive integers, describing the coefficients for reactants and products involved at each elementary reaction step [28]. In this notation, *nr* stands for the number of reactions. The matrix of stoichiometric coefficients denoted **N** (*nr* x *ns*), containing negative coefficients for reactants and positive ones for products, can be calculated from the difference **P – E**. Matrices **E**, **P** and **N** are illustrated in Figure 4.1 for an example mechanism, thoroughly discussed in Section 4.4.2.

$$A + B \xrightarrow{k_1} C$$
$$A + C \xrightarrow{k_2} D$$
$$nr = 2 \text{ and } ns = 4 \ (A, B, C, D)$$
$$\mathbf{E}(2 \text{ x } 4) = \begin{matrix} \\ \end{matrix} \begin{matrix} A & B & C & D \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{P}(2 \text{ x } 4) = \begin{matrix} \\ \end{matrix} \begin{matrix} A & B & C & D \end{matrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{N}(2 \text{ x } 4) = \mathbf{P} - \mathbf{E} = \begin{matrix} \\ \end{matrix} \begin{matrix} A & B & C & D \end{matrix} \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 \end{bmatrix}$$

Figure 4.1: Matrices **E**, **P** and **N** for a mechanism including *nr* = 2 reactions and *ns* = 4 species.

Using initial and dosing conditions, as well as the matrices **E** and **N**, the matrix of concentration profiles **C** can be calculated for a given set of rate constants by numerical integration of the corresponding system of ordinary differential equations describing the kinetic model. We refer to [14,29] for details about the integration of these differential equations expressed in concentration terms.

However, expressed in concentration terms, these *ns* differential equations often comprise redundant terms, as several of these equations can be expressed as a linear combination of others. The number of differential equations to integrate can be substantially reduced if the matrix of extents of reaction **X** (*nt* x *nr*) is used instead of **C** [30]. The extents of reaction **X**, expressed in concentration units, can be defined as the integrated rate of each reaction step at time *t*, i.e. the amount consumed or generated by each reaction normalised to unity stoichiometry [31]. An equation to calculate **C** from **X** can be deduced from the definition of **X** under batch conditions.

$$\mathbf{C} = \mathbf{X}\mathbf{N} + \mathbf{C_0} \tag{4.2}$$

The matrix of initial concentrations $\mathbf{C_0}$ (*nt* x *ns*) is obtained by *nt*-times identical vertical stacking of the row vector of initial concentrations $\mathbf{c_0}$ (1 x *ns*). This stacking operation can be formally performed by left multiplying $\mathbf{c_0}$ with a column vector of ones, denoted **1**, of dimensions (*nt* x 1).

$$\mathbf{C_0} = \mathbf{1}\mathbf{c_0} \tag{4.3}$$

Performing a mass balance on the kinetic process, Equation 4.2 can be extended to also cover semi-batch conditions leading to an expression that is identical to the one proposed by Amrhein et al., known as the factorisation of concentration data [27].

$$\mathbf{C} = \mathbf{X}\mathbf{N} + \mathbf{D}\mathbf{C_0} + \left(\mathbf{I} - \mathbf{D}\right)\mathbf{C_{dos}} \tag{4.4}$$

where $\mathbf{C_{dos}}$ (*nt* x *ns*) denotes the matrix of dosing concentrations for all species at any time.

It is calculated from a matrix $\mathbf{C_{in}}$ (*nf* x *ns*) comprising the dosing concentrations associated with each dosing step. The row dimension is indicated by *nf*, the number of dosing steps. $\mathbf{C_{dos}}$ is generated by identical vertical stacking of each row of $\mathbf{C_{in}}$ over the time domain corresponding to the start time and the end time of each single dosing step. $\mathbf{C_{dos}}$ contains zeros at times (rows) when no species is dosed. The stacking procedure leading to $\mathbf{C_{dos}}$ is visualised in Figure 4.2.

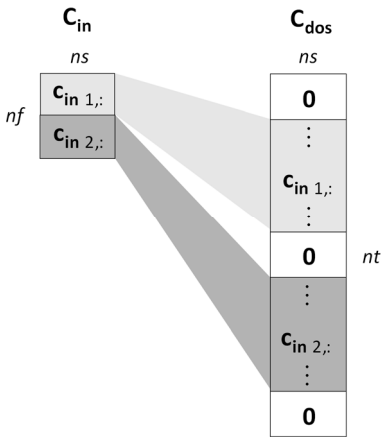

Figure 4.2: Setup of $\mathbf{C_{dos}}$ (*nt* x *ns*) by identical vertical stacking of each row of $\mathbf{C_{in}}$ (*nf* x *ns*).

In Equation 4.4, **I** denotes the identity matrix of dimensions (*nt* x *nt*), and **D** (*nt* x *nt*) the diagonal matrix of dilution calculated from the initial volume v_0 and the vector of time resolved reaction volumes **v** (*nt* x 1). Note that v_0 is also v_1, i.e. the first element of the vector **v**.

$$\mathbf{D} = v_0 \mathrm{DIAG}\left(\mathbf{v}\right)^{-1} \tag{4.5}$$

The DIAG operator generates a diagonal matrix from the corresponding vector argument [32]. Note that in batch conditions (assuming constant density during the reaction), **D** = **I** and Equation 4.4 collapses to Equation 4.2.

The matrix of extents of reaction, **X**, can be calculated by integrating nr differential equations, with initial conditions $\mathbf{x_0}$ (1 x *nr*) = **0**. However, since kinetic rate laws are intrinsically defined in concentration terms (not considering activity coefficients), calculating the derivative of **X** at each integration step *t* requires the concentrations **C** to be recalculated from **X**. This procedure is described in Equations 4.6 to 4.8 for a constant density.

$$\frac{\mathrm{d}x_{\mathrm{t,j}}}{\mathrm{d}t} = k_{\mathrm{j}} \prod_{i=1}^{ns} c_{\mathrm{t,i}}^{e_{\mathrm{j,i}}} - \frac{f_{\mathrm{t}}}{v_{\mathrm{t}}} x_{\mathrm{t,j}} \qquad \text{for } j = 1 \text{ to } nr \tag{4.6}$$

$$\frac{\mathrm{d}v_{\mathrm{t}}}{\mathrm{d}t} = f_{\mathrm{t}} \tag{4.7}$$

With

$$c_{\mathrm{t,i}} = \prod_{j=1}^{nr} x_{\mathrm{t,j}} n_{\mathrm{j,i}} + d_{\mathrm{t,t}} c_{0\ \mathrm{t,i}} + \left(1 - d_{\mathrm{t,t}}\right) c_{\mathrm{dos\ t,i}} \qquad \text{for } i = 1 \text{ to } ns \text{ and } d_{\mathrm{t,t}} = \frac{v_0}{v_{\mathrm{t}}} \tag{4.8}$$

and f_{t}, the element of the vector of dosing rates **f** (*nt* x 1) at time *t*, and $e_{\mathrm{j,i}}$ being the corresponding reactant coefficients (elements of **E**). For each time *t*, $c_{\mathrm{t,i}}$ is calculated according to Equation 4.8 (see also Equation 4.4) and substituted into Equation 4.6 to allow integrating the system of differential equations described by Equations 4.6 and 4.7.

Note, if a mechanism is comprised by reversible reactions, matrix **N** can be simplified by keeping only the rows corresponding to the forward reactions and **C** can be calculated by computing the net forward extent of reaction defined as the difference between the forward and backward extents for each single reversible reaction.

In the method proposed by Maeder and Zuberbühler for the kinetic hard-modelling of spectroscopic data, the product of the integrated concentration profiles **C** and the pure component spectra **A** is compared to the measured data matrix **Y**, and results in the residuals **R** = **Y – CA**, which capture the differences between the measured and the modelled absorbances.

In least-squares analysis, the sum of all squared residuals **R** is used as the objective function to be minimised by iteratively optimising the non-linear kinetic parameters, i.e. the rate constants defining the concentration profiles in **C**. As **A** is comprised of linear parameters only, it can be eliminated in each iteration from the non-linear optimisation and can be replaced by its linear least-squares estimate [10,14].

$$\mathbf{A} = \mathbf{C}^{+}\mathbf{Y} = \left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{Y} \tag{4.9}$$

Where $\mathbf{C}^{+}$ denotes the Moore-Penrose left pseudo-inverse of **C**.

4.3.1 Rank deficiencies in the concentration matrix C

A frequent problem in the method of Maeder and Zuberbühler for the kinetic hard-modelling of spectroscopic data is encountered when the matrix of concentration profiles **C** (*nt* x *ns*) is rank deficient and thus $\mathbf{C}^{+}$ cannot be computed. Consequently, **A** cannot be eliminated from the non-linear optimisation using Equation 4.9, except if one of the strategies mentioned in Table 4.1 is used. In order to understand the origin of possible rank deficiencies in **C** and to elaborate a systematic method to select the appropriate strategy, the kernel of **C** and some other related concepts are briefly defined.

4.3.1.1 Kernel of C

The dimension of **C** (*nt* x *ns*), denoted dim(**C**), is the minimum number of coordinates required to define a point in the space spanned by the columns or the rows of **C**. Thus, the dimension is the smaller of the number of rows (*nt*) or columns (*ns*). As in kinetic experiments *nt* is generally greater than *ns*, the dimension of **C** is *ns*.

$$\dim\left(\mathbf{C}\right) = ns \tag{4.10}$$

The image of **C**, denoted im **C**, is the vector space spanned by the linearly independent columns of **C**, which form a vector basis. The (chemical) rank of **C**, denoted rank(**C**) or dim(im **C**), is the dimension of its image and is used to calculate the number of species with linearly independent concentration profiles, *nc*.

$$nc = \text{rank}(\mathbf{C}) \tag{4.11}$$

The kernel of **C**, denoted ker **C**, is the vector space of dimension dim(ker **C**) spanned by the vectors forming the null space **0** when left-multiplied by **C**, i.e. **C** (ker **C**) = **0** [33]; this can also be seen as the mass balance (or closure). Note that the kernel is a base that defines a vector space, and thus any linear combinations of the kernel also form a basis for this vector space (rotational ambiguity, see Equation 4.13). As a vector base, the kernel always contains the trivial solution (**0**). The kernel of **C** describes the linear dependencies between the ns concentration profiles (columns of **C**) and thus a row in the kernel only comprised by zeros indicates that the corresponding species is linearly independent from the others. The kernel of **C** can be calculated from its reduced row echelon form (RREF). However, for stability reasons, Singular Value Decomposition (SVD) is preferred. In Matlab, an orthonormal basis for the kernel of **C** can be computed with the null command [34]; it performs SVD and returns the (*ns*–*nc*) eigenvectors of $\mathbf{C}^T\mathbf{C}$, for which the corresponding eigenvalues are zero within the numerical accuracy. The null command can also be used within the symbolic toolbox of Matlab in order to obtain an analytical expression for the kernel. As will be shown in Section 4.4, this can have an advantage to describe the effect of different experimental conditions.

The rank-nullity theorem states that the dimension of the kernel plus the dimension of the image of **C** equals the dimension of **C** [33], i.e. equals *ns* under the condition that $nt \geq ns$. A rank deficiency in **C** occurs if *nc* is lower than *ns*, i.e. if the kernel of **C** is comprised of more than the trivial solution (**0**). The dimension of the kernel of **C** is also referred to as the defect of **C**.

$$nc + \dim(\ker \mathbf{C}) = ns \tag{4.12}$$

Finally, we define the property of pseudo-equivalence denoted $\sim$ of two matrices of different row but identical column dimension that share the same kernel. The terminology ‘pseudo’ has been introduced as mathematical equivalence would require identical matrix dimensions.

As an example, the property of pseudo-equivalence between **C** and any matrix $\boldsymbol{\Omega}$ implies:

$$\mathbf{C} \sim \boldsymbol{\Omega} \;\Leftrightarrow\; (\ker \mathbf{C})\mathbf{T}_{\mathbf{lin}} = \ker\, \boldsymbol{\Omega} \tag{4.13}$$

Where $\mathbf{T}_{\mathbf{lin}}$ is a linear transformation, i.e. a square matrix of dimensions (*ns* x *ns*), applied on ker **C** to account for the rotational ambiguity of ker $\boldsymbol{\Omega}$, as mentioned before.

4.3.1.2 A pseudo-equivalent matrix for C under semi-batch conditions

The prediction of the rank of **C** has been discussed in literature [15,20,21,25,26] and requires defining the number of independent reactions [15,21]. According to Amrhein et al, the number of independent reactions, here denoted nr_i, is the number of reactions that have (i) constant and linearly independent stoichiometric coefficients (rows of **N**), and (ii) linearly independent rate laws (columns of **X**). The number of independent reactions, which is also the minimum number of differential equations to be integrated, can be used to determine the rank of matrix **C**.

$$\operatorname{rank}(\mathbf{C}) = \min(nr_i + 1,\ ns) \tag{4.14}$$

Identifying the independent reactions can be a difficult task when the kinetic model is complex. Thus, here we propose to analyse the linear dependencies in **C** based on the full kinetic model introducing a pseudo-equivalent matrix, $\boldsymbol{\Omega}$, that covers possible rank deficiencies in **X** (generally due to parallel reactions) and in **N** (generally due to reversible reactions). Based on Equation 4.4, an augmented matrix $\boldsymbol{\Omega}$ (3·*nt* x *ns*) can be built by vertically stacking matrices **XN**, $\mathbf{DC_0}$ and $(\mathbf{I} - \mathbf{D})\mathbf{C_{dos}}$, as shown in Equation 4.15.

$$\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{XN} \\ \hline \mathbf{DC_0} \\ \hline (\mathbf{I}-\mathbf{D})\mathbf{C_{dos}} \end{bmatrix} \sim \mathbf{C} \quad (\text{under semi-batch conditions}) \tag{4.15}$$

This time variant matrix $\boldsymbol{\Omega}$ (3·*nt* x *ns*) can be reduced to a time invariant matrix $\boldsymbol{\Omega}$ (*ns*+*nf*+1 x *ns*) (see Section 4.6.1 – Appendix) so that the system of differential equations no longer needs to be integrated. For the sake of simplicity, the same matrix symbol ($\boldsymbol{\Omega}$) is used in Equations 4.15 and 4.16.

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}\left(\mathbf{k}\right)\mathbf{N}}{\mathbf{c_0}} \\ \hline \mathbf{C_{in}} \end{bmatrix} \sim \mathbf{C} \quad \left(\text{under semi-batch conditions}\right) \tag{4.16}$$

With μ being an arbitrary positive scalar different from 0 or 1, matrix **1** (*ns* x *nr*) comprised of ones only and the superscript $\bullet\mathbf{E}^T$ representing the element-wise raise to the power of $\mathbf{E}^T$.

If vector **k** is unknown, e.g. the rate constants are to be optimised, it can be omitted in the calculation of $\mathbf{\Omega}$, i.e. **k** = **1**, leading to a simplified augmented matrix that is not strictly pseudo-equivalent to **C**. However, the rank of this simplified matrix is still identical to the rank of **C** and its kernel has the identical rows comprised by zeros only as has the kernel of **C**. Note that strict pseudo-equivalence between $\mathbf{\Omega}$ and **C** (and the inclusion of DIAG(**k**) in Equation 4.16) is required for an explicit determination of the relationship between the appearance of zero entries in ker **C** and the stoichiometric conditions. A discussion of ker **C** depending on stoichiometric conditions is given in Examples 1 to 3 (see Section 4.4) using the symbolic toolbox of Matlab [34].

4.3.2 A systematic method for selecting the appropriate strategy to treat rank deficiencies in C

Based on the analysis of the kernel of matrix $\mathbf{\Omega}$, a systematic method is developed for selecting the appropriate strategy to treat rank deficiencies in **C**, i.e. to select which absorbing species to set uncoloured (Strategy 1), which known pure component spectra to provide (Strategy 2), which species to dose (Strategy 3), or which initial concentrations to vary (Strategy 4). As dosing is just one strategy amongst the four presented here, the proposed method is initialised with an augmented matrix $\mathbf{\Omega}$ of dimensions (*ns*+1 x *ns*) comprised of the first two segments in Equation 4.16. The third segment ($\mathbf{C_{in}}$) is included only if dosing is performed, i.e. if Strategy (3) is applied.

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}\left(\mathbf{k}\right)\mathbf{N}}{\mathbf{c_0}} \end{bmatrix} \sim \mathbf{C} \quad \left(\text{under batch conditions}\right) \tag{4.17}$$

The method relies on the following seven assumptions: (i) the kinetic model is correct; (ii) no reactions are identical, i.e. no two reactions share the same rate law and rate constants, and form the same products; (iii) all species required to initiate the reactions are present initially or are dosed (Strategy 3); (iv) the true pure component spectra of the absorbing species are linearly independent; (v) each true pure component spectrum provided in Strategy (2) is correct, e.g. no baseline shift or the baseline shift is corrected by an appropriate method [16,35]; (vi) all species absorb or the truly non-absorbing species are known and Strategy (2) can be applied (a pure component spectrum comprised by zeros only is provided); (vii) the time resolution in **Y** is such that all reaction steps defined by the kinetic model are covered.

In the following, we describe the different steps required for our systematic approach as outlined in Figure 4.3 and Figure 4.4.

Note that rank deficiencies in the kinetic hard-modelling of calorimetric data have also been investigated from a theoretical point of view. These rank deficiencies come from linear dependencies in the reaction extents **X** and only appear in case of two identical rows in matrix **E**, i.e. for strictly parallel reactions. This causes an analogue problem as in spectroscopy, when pure component spectra are fitted by Equation 4.9, in the sense that, in calorimetry, a pseudo-inverse matrix cannot be computed during the fitting of the reaction enthalpies. Linear dependencies can be predicted based on matrix **E** and effects of applying Strategy (1) or (2) can be predicted without the need of introducing a time invariant matrix $\boldsymbol{\Omega}$. For more details about linear dependencies in kinetic hard-modelling of calorimetric data, see Section 4.6.5 (Appendix).

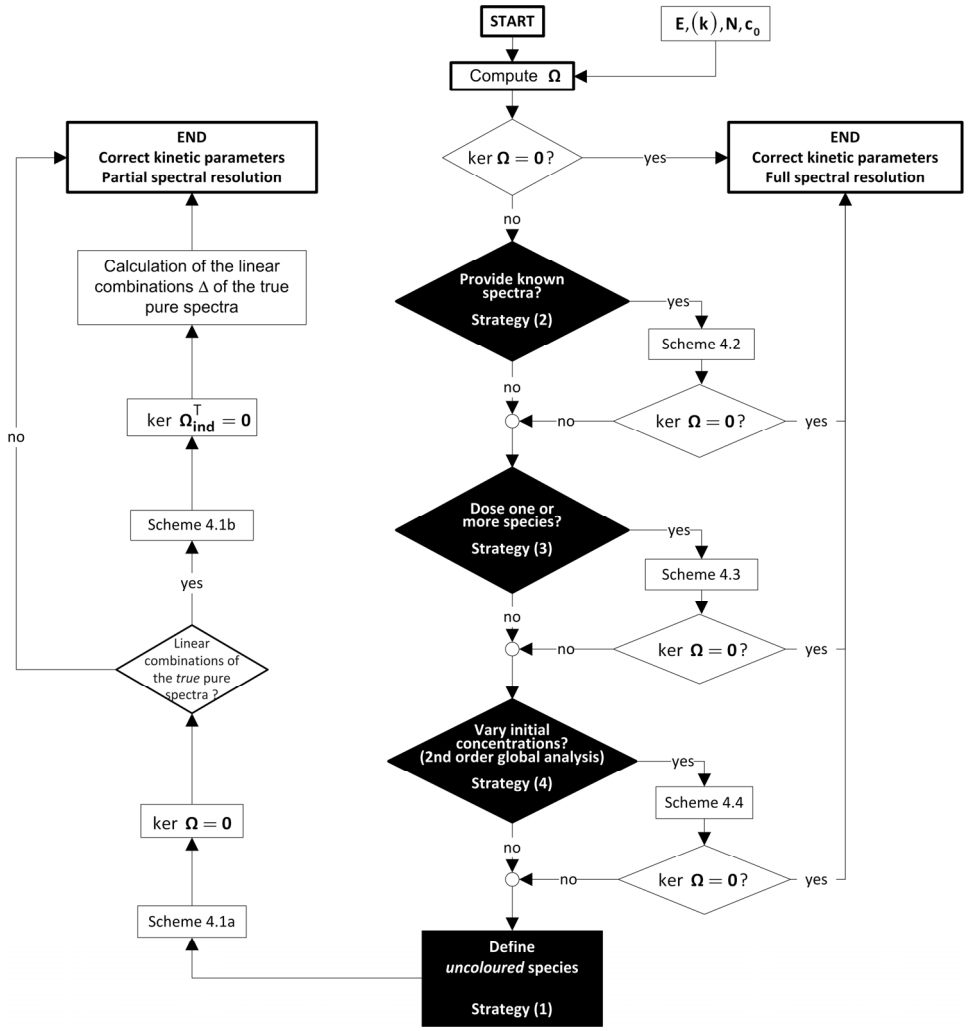

Figure 4.3: Description of the different steps involved in the proposed systematic method for selecting the appropriate strategy to treat rank deficiencies in the matrix of concentration profiles.

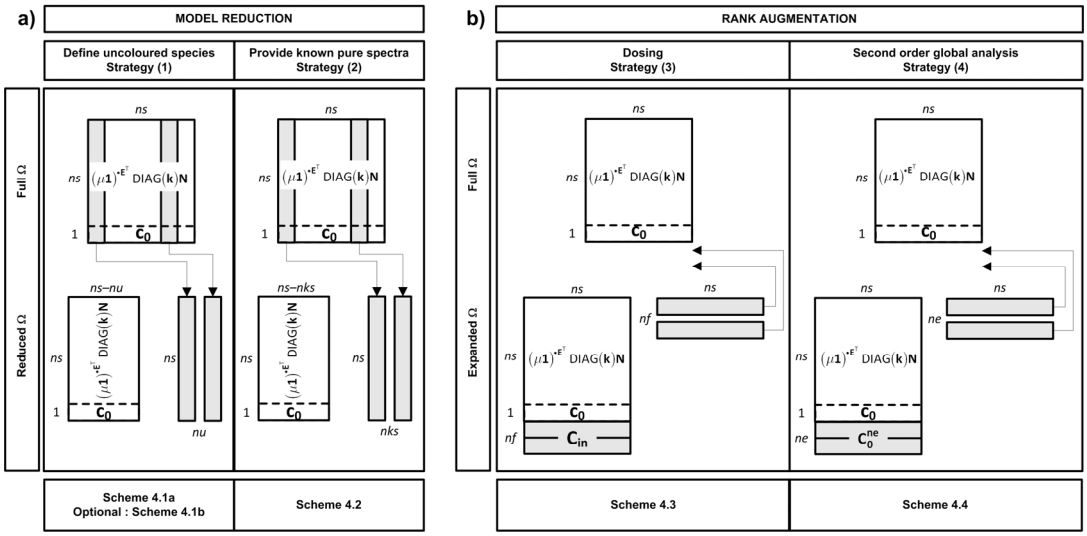

Figure 4.4: (a) Reduction of matrix **Ω** when strategies (1) and (2) are applied (shaded columns are removed) and (b) expansion of **Ω** when strategies (3) and (4) are used (shaded rows are appended).

4.3.2.1 Prediction of the uncoloured species to include in Strategy (1)

This strategy is a mathematical treatment to reduce the matrix of concentration profiles **C** such that its pseudo-inverse is defined (see Equation 4.9), while keeping the same least-squares optimum for the rate constants. It consists in reducing Beer's law (Equation 4.1) by defining a set of (*ns*–*nu*) coloured and *nu* uncoloured species, and by eliminating from the matrix **C** (*nt* x *ns*) the columns corresponding to these *nu* uncoloured species, thus keeping only linearly independent concentration profiles. In this context, we define a species as uncoloured if it is a true absorber but set to non-absorbing in order to avoid a rank deficiency in **C**. Note that if a species is known a priori to be truly non-absorbing, it should not be defined uncoloured but rather treated by Strategy (2), providing a known pure component spectrum containing zeros only. This reduction in the spectroscopic model leads to a concentration matrix, denoted $\mathbf{C_c}$ (*nt* x *ns*–*nu*), comprised of (*ns*–*nu*) coloured species only. This affects the rank-nullity theorem, as the dimension of the image plus the dimension of the kernel of $\mathbf{C_c}$ equals now the reduced number of species, i.e. (*ns*–*nu*).

$$nc + \dim\left(\ker \mathbf{C_c}\right) = ns - nu \quad , \quad \left(\ker \mathbf{C_c}\right)\mathbf{T_{lin}} = \ker\ \mathbf{\Omega} \tag{4.18}$$

Where $\mathbf{T_{lin}}$ is a linear transformation, i.e. a square matrix of dimensions (*ns*–*nu* x *ns*–*nu*), applied on ker $\mathbf{C_c}$ to account for the rotational ambiguity of ker $\mathbf{\Omega}$.

A matrix $\mathbf{\Omega}$ of dimensions (*ns*+1 x *ns*–*nu*) that is pseudo-equivalent to $\mathbf{C_c}$ (*nt* x *ns*–*nu*) can be obtained by eliminating from the original augmented matrix $\mathbf{\Omega}$ (*ns*+1 x *ns*) the columns corresponding to the *nu* uncoloured species in the same way **C** (*nt* x *ns*) is reduced to $\mathbf{C_c}$ (*nt* x *ns*–*nu*). This elimination is presented in Figure 4.4a (Strategy 1). The appropriate set of uncoloured species to be defined in order to allow the computation of the pseudo-inverse of $\mathbf{C_c}$ can be predicted iteratively (see Scheme 4.1a) based on the non-zero rows of ker $\mathbf{\Omega}$.

Defining uncoloured species, although they are in fact contributing to the spectroscopic matrix **Y**, has the drawback that the missing spectral contributions of the uncoloured species are linearly transferred into the pure component spectra of the coloured species. Hence, the fitted pure component spectra of the coloured species calculated from Equation 4.9 are comprised of linear combinations of the true (resolved) pure component spectra, and thus are often difficult to interpret. Since no spectral resolution can be achieved with this strategy, it should only be employed when no other strategy (2, 3 or 4) can be applied satisfactory.

Scheme 4.1a: Iterative method for selecting appropriate uncoloured species in Strategy (1) in order to enable the calculation of the pseudo-inverse of the concentration matrix.

Step	Action
Initial assumption	Initially, $\mathbf{\Omega}$ has dimensions (*ns*+1 x *ns*) [a]
	All species are set coloured
	ker $\mathbf{\Omega}$ is calculated
Iterative step 1	The uncoloured species is chosen amongst the non-zero rows of ker $\mathbf{\Omega}$
	The column corresponding to the chosen species is removed from $\mathbf{\Omega}$
Iterative step 2	ker $\mathbf{\Omega}$ is recalculated
Stop criterion	Iterative steps 1–2 are repeated until ker $\mathbf{\Omega}$ = **0**
	Finally, $\mathbf{\Omega}$ has dimensions (*ns*+1 x *ns*–*nu*) [b]

[a] Initially $\mathbf{\Omega}$ has dimensions (*ns*+1 x *ns*–*nks*), (*ns*+*nf*+1 x *ns*) or (*ns*+*ne*+1 x *ns*) respectively if Strategies (2), (3) or (4) were applied before Strategy (1).

[b] Finally $\mathbf{\Omega}$ has dimensions (*ns*+1 x *ns*–*nu*–*nks*), (*ns*+*nf*+1 x *ns*–*nu*) or (*ns*+*ne*+1 x *ns*–*nu*) respectively if Strategies (2), (3) or (4) were applied before Strategy (1).

Optionally, the underlying linear combinations of the true pure component spectra can be determined by eliminating the linear dependencies in the rows of the augmented matrix $\mathbf{\Omega}$ (*ns*+1 x *ns*). This elimination, leading to a matrix $\mathbf{\Omega_{ind}}$ of dimensions (nr_i+1 x *ns*), is required as $\mathbf{\Omega}$ uses the full kinetic model to cover possible simultaneous rank deficiencies in **X** and **N**.

The appropriate set of linearly independent rows in $\mathbf{\Omega}$ can be predicted iteratively based on the non-zero rows of ker $\mathbf{\Omega}^T$ (see Scheme 4.1b). From this, the number of linearly independent reactions nr_i (see Equation 4.14) can also easily be determined.

$$nr_i = ns - \dim\left(\ker\ \mathbf{\Omega}^T\right) \tag{4.19}$$

Once coloured species and linearly independent reactions have been identified, matrix $\mathbf{\Delta}$ (*ns*–*nu* x *ns*) comprising the coefficients for the linear combinations of the true (resolved) pure component spectra can be calculated using Equation 4.20.

$$\mathbf{\Delta} = \mathbf{C}_\mathbf{c}^+\mathbf{C} = \left(\mathbf{\Omega}\Big|_{\substack{\text{comprised of}\\\text{coloured species}}}\right)^+ \mathbf{\Omega}\Big|_{\substack{\text{comprised of}\\\text{all species}}} = \left(\mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\\text{coloured species}}}\right)^{-1} \mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\\text{all species}}} \tag{4.20}$$

With $\mathbf{\Omega}\Big|_{\substack{\text{comprised of}\\\text{coloured species}}}$, $\mathbf{\Omega}\Big|_{\substack{\text{comprised of}\\\text{all species}}}$, $\mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\\text{coloured species}}}$ and $\mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\\text{all species}}}$ having dimensions (*ns*+1 x *ns*–*nu*), (*ns*+1 x *ns*), (nr_i+1 x *ns*–*nu*) and (nr_i+1 x *ns*) respectively, and with nr_i+1=*ns*–*nu*=*nc*.

Note that a strictly pseudo-equivalent $\boldsymbol{\Omega}$ is required in Equation 4.20, i.e. the vector **k** must be comprised of correct known rate constants or has to be defined symbolically.

Scheme 4.1b: Optional iterative method for selecting appropriate linearly independent rows in Strategy (1) in order to calculate the linear combinations of the true (resolved) pure component spectra.

Step	Action
Initial assumption	Initially, $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*) [a)] All rows are linearly independent ker $\boldsymbol{\Omega}^T$ is calculated
Iterative step 1	The linearly dependent row is chosen amongst the non-zero rows of ker $\boldsymbol{\Omega}^T$ The corresponding row is removed from $\boldsymbol{\Omega}$
Iterative step 2	ker $\boldsymbol{\Omega}^T$ is recalculated
Stop criterion	Iterative steps 1–2 are repeated until ker $\boldsymbol{\Omega}^T = \mathbf{0}$ The resulting matrix $\boldsymbol{\Omega}$ is denoted $\boldsymbol{\Omega}_{\mathbf{ind}}$ Finally, $\boldsymbol{\Omega}$ has dimensions (nr_i+1 x *ns*) [b)]

a) Initially $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*–*nks*), (*ns*+*nf*+1 x *ns*) or (*ns*+*ne*+1 x *ns*) respectively if Strategies (2), (3) or (4) were applied before Strategy (1).

b) Finally $\boldsymbol{\Omega}$ has dimensions (nr_i+1 x *ns*–*nks*), (nr_i+*nf*+1 x *ns*) or (nr_i+*ne*+1 x *ns*) respectively if Strategies (2), (3) or (4) were applied before Strategy (1).

The linear combinations $\boldsymbol{\Delta}$ obtained by Equation 4.20 are the coefficients by which the true pure component spectra **A** have to be weighted to obtain component spectra $\mathbf{A_c}$ (*nc* x *nw*) of the coloured species fitted by Equation 4.9.

$$\mathbf{A_c} = \boldsymbol{\Delta}\,\mathbf{A} = \mathbf{C_c^+ Y} \tag{4.21}$$

Note that if Strategy (1) is combined with Strategy (2) providing *nks* pure component spectra, **A** (*ns* x *nw*) is replaced in Equation 4.21 by the (*ns*–*nks*) remaining true pure component spectra, and that *nks* columns, corresponding to the species of known spectra, have to be removed from $\boldsymbol{\Omega}$. For details about Strategy (2), see Section 4.3.2.2.

4.3.2.2 Prediction of the known pure spectra to provide in Strategy (2)

This strategy consists in separating **Y** (*nt* x *nw*) into a known contribution due to *nks* provided pure component spectra and into an unknown contribution due to the (*ns*–*nks*) unknown pure component spectra. The known contributions, i.e. the product of the provided pure component spectra and the corresponding concentration profiles, are subtracted from the overall spectroscopic data matrix **Y**. The remaining concentration profiles corresponding to the species of unknown spectra are collected in a reduced matrix, denoted $\mathbf{C}_{\mathbf{uk}}$ (*nt* x *ns*–*nks*). See [10,19] for more details about the subtraction of the known contribution from the overall absorbance during the kinetic fitting process.

This reduction of the spectroscopic model (Beer's law) also affects the rank-nullity theorem, as the dimension of the image plus the dimension of the kernel of $\mathbf{C}_{\mathbf{uk}}$ must now be equivalent to the reduced number of species, i.e. (*ns*–*nks*).

$$nc + \dim\left(\ker \mathbf{C}_{\mathbf{uk}}\right) = ns - nks \;\; , \;\; \left(\ker \mathbf{C}_{\mathbf{uk}}\right)\mathbf{T}_{\mathbf{lin}} = \ker \boldsymbol{\Omega} \tag{4.22}$$

Where $\mathbf{T}_{\mathbf{lin}}$ is a square matrix of dimensions (*ns*–*nks* x *ns*–*nks*) applying a linear transformation on ker $\mathbf{C}_{\mathbf{uk}}$ to account for the rotational ambiguity of ker $\boldsymbol{\Omega}$.

When $\mathbf{C}_{\mathbf{uk}}$ is of full rank, the unknown pure component spectra $\mathbf{A}_{\mathbf{uk}}$ (*ns*–*nks* x *nw*), i.e. the ones that have not been provided, can be linearly fitted in analogy to Equation 4.9. A pseudo-equivalent matrix for $\mathbf{C}_{\mathbf{uk}}$ can be obtained from a corresponding reduction of $\boldsymbol{\Omega}$ (*ns*+1 x *ns*) to $\boldsymbol{\Omega}$ (*ns*+1 x *ns*–*nks*) following the same procedure as the reduction of **C** (*nt* x *ns*) to $\mathbf{C}_{\mathbf{uk}}$ (*nt* x *ns*–*nks*). This procedure is presented in Figure 4.4a (Strategy 2).

The appropriate list of species for which pure component spectra should be provided in order to fully resolve the remaining (*ns*–*nks*) fitted pure component spectra, can be predicted iteratively (see Scheme 4.2) based on the non-zero rows of the kernel of $\boldsymbol{\Omega}$ (*ns*+1 x *ns*–*nks*). Note that iterative Scheme 4.2 can always be stopped at any point before reaching ker $\boldsymbol{\Omega}$ = **0** and in such case, Strategy (2) has to be combined with another strategy of Table 4.1.

Scheme 4.2: Iterative method for selecting appropriate species in Strategy (2) to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*) [a)] No pure component spectrum is provided ker $\boldsymbol{\Omega}$ is calculated
Iterative step 1	The species for which a pure component spectrum has to be provided is chosen amongst the non-zero rows of ker $\boldsymbol{\Omega}$ The column corresponding to the chosen species is removed from $\boldsymbol{\Omega}$
Iterative step 2	ker $\boldsymbol{\Omega}$ is recalculated
Stop criterion	Iterative steps 1–2 are repeated until ker $\boldsymbol{\Omega}$ = **0** Finally, $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*–*nks*) [b)]

a) Initially $\boldsymbol{\Omega}$ has dimensions (*ns*+*nf*+1 x *ns*) or (*ns*+*ne*+1 x *ns*) respectively if Strategies (3) or (4) were applied before Strategy (2).

b) Finally $\boldsymbol{\Omega}$ has dimensions (*ns*+*nf*+1 x *ns*–*nks*) or (*ns*+*ne*+1 x *ns*–*nks*) respectively if Strategies (3) or (4) were applied before Strategy (2).

4.3.2.3 Prediction of the species to dose in Strategy (3)

In this strategy, one or more species are dosed in a sequence of *nf* dosing steps in order to augment the rank of **C** to its full rank. A matrix $\boldsymbol{\Omega}$ of dimensions (*ns*+*nf*+1 x *ns*) that is pseudo-equivalent to **C** can be obtained by expanding $\boldsymbol{\Omega}$ (*ns*+1 x *ns*) with the appropriate rows corresponding to the dosed species as presented in Equation 4.16. The expansion of $\boldsymbol{\Omega}$ (*ns*+1 x *ns*) to $\boldsymbol{\Omega}$ (*ns*+*nf*+1 x *ns*) is also outlined in Figure 4.4b (Strategy 3).

Scheme 4.3: Iterative method for selecting appropriate dosing species in Strategy (3) to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*) [a)] No species is dosed ker $\boldsymbol{\Omega}$ is calculated
Iterative step 1	The species to be dosed is chosen amongst the non-zero rows of ker $\boldsymbol{\Omega}$ The dosed species adds a row in $\mathbf{C}_{in}$ and thus in $\boldsymbol{\Omega}$
Iterative step 2	ker $\boldsymbol{\Omega}$ is recalculated
Stop criterion	Iterative steps 1–2 are repeated until ker $\boldsymbol{\Omega}$ = **0** Finally, $\boldsymbol{\Omega}$ has dimensions (*ns*+*nf*+1 x *ns*) [b)]

a) Initially $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*–*nks*) or (*ns*+*ne*+1 x *ns*) respectively if Strategies (2) or (4) were applied before Strategy (3).

b) Finally $\boldsymbol{\Omega}$ has dimensions (*ns*+*nf*+1 x *ns*–*nks*) or (*ns*+*nf*+*ne*+1 x *ns*) respectively if Strategies (2) or (4) were applied before Strategy (3).

In order to fully resolve the fitted pure component spectra, the appropriate set of species to dose can be predicted iteratively using the non-zeros rows of ker $\mathbf{\Omega}$ (see Scheme 4.3). Note that iterative Scheme 4.3 can always be stopped before reaching ker $\mathbf{\Omega} = \mathbf{0}$ and in such case, Strategy (3) has to be combined with another strategy of Table 4.1.

4.3.2.4 Prediction of the initial concentrations to vary in Strategy (4)

In this strategy, *ne* additional kinetic experiments ($\mathbf{Y_i}$ with i = 2, …, *ne*+1) are performed under different initial concentrations ($\mathbf{c}_0^{\exp_i}$), appended vertically to the first experiment and subsequently analysed together by second order global analysis [10,14,22-24]. Each individual spectroscopic measurement $\mathbf{Y_i}$ and each modelled associated concentration matrix $\mathbf{C_i}$ is concatenated vertically to form matrices $\mathbf{Y_{glob}}$ and $\mathbf{C_{glob}}$ respectively, replacing **Y** and **C** in Beer's law (Equation 4.1). If the concatenated matrix of concentrations $\mathbf{C_{glob}}$ is of full rank and if one unique set of pure component spectra is assumed for all experiments (global mode), the pseudo-inverse of $\mathbf{C_{glob}}$ can be calculated in analogy to Equation 4.9 and leads to fully resolved pure component spectra.

A matrix $\mathbf{\Omega}$ ((*ne*+1)·*ns*+*ne*+1 x *ns*) that is pseudo-equivalent to $\mathbf{C_{glob}}$ could be obtained by concatenating vertically as many augmented matrices $\mathbf{\Omega_i}$ with different initial concentrations as the number of concentration matrices $\mathbf{C_i}$ concatenated in $\mathbf{C_{glob}}$. In global mode, however, this vertical augmentation is only possible if one common kinetic model is assumed between all individual experiments, i.e. the rate laws and the matrices of stoichiometry are identical between the (*ne*+1) experiments. This common property between all the experiments allows reducing the dimensionality of the matrix $\mathbf{\Omega}$ to (*ns*+*ne*+1 x *ns*) by keeping only once the product $\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^T}$ DIAG$\left(\mathbf{k}\right)\mathbf{N}$.

Therefore, a pseudo-equivalent matrix for $\mathbf{C_{glob}}$ can be obtained by augmenting just the initial $\mathbf{\Omega}$ (*ns*+1 x *ns*) with *ne* rows corresponding to the different vectors of initial concentrations of the additional experiments, collected in a matrix $\mathbf{c}_0^{ne}$ (*ne* x *ns*), as shown in Figure 4.4b (Strategy 4).

$$\mathbf{C}_0^{ne} = \begin{bmatrix} \mathbf{c}_0^{exp_2} \\ \hline \mathbf{c}_0^{exp_3} \\ \hline \vdots \\ \hline \mathbf{c}_0^{exp_{ne+1}} \end{bmatrix} \tag{4.23}$$

The appropriate list of species for which the initial concentration should be varied in order to fully resolve the fitted pure component spectra can be predicted iteratively using the non-zeros rows of ker $\boldsymbol{\Omega}$ (see Scheme 4.4). Note that iterative Scheme 4.4 can always be stopped before reaching ker $\boldsymbol{\Omega}$ = $\mathbf{0}$ and in such case, Strategy (4) has to be combined with another strategy of Table 4.1.

Scheme 4.4: Iterative method for selecting appropriate species for which the initial concentration should be varied in Strategy (4) in order to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*) [a]
	Only one experiment with initial concentrations $\mathbf{c_0}$ (1 x *ns*) is considered
	ker $\boldsymbol{\Omega}$ is calculated
Iterative step 1	The species whose initial concentration should be varied is chosen amongst the non-zero rows of ker $\boldsymbol{\Omega}$
	The new set of initial concentrations adds a row $\mathbf{C}_0^{ne}$ and thus in $\boldsymbol{\Omega}$
Iterative step 2	ker $\boldsymbol{\Omega}$ is recalculated
Stop criterion	Iterative steps 1–2 are repeated until ker $\boldsymbol{\Omega}$ = $\mathbf{0}$
	Finally, $\boldsymbol{\Omega}$ has dimensions (*ns*+*ne*+1 x *ns*) [b]

[a] Initially $\boldsymbol{\Omega}$ has dimensions (*ns*+1 x *ns*–*nks*) or (*ns*+*nf*+1 x *ns*) respectively if Strategies (2) or (3) were applied before Strategy (4).

[b] Finally $\boldsymbol{\Omega}$ has dimensions (*ns*+*ne*+1 x *ns*–*nks*) or (*ns*+*nf*+*ne*+1 x *ns*) respectively if Strategies (2) or (3) were applied before Strategy (4).

4.4 Examples

The systematic method for selecting the appropriate strategy to treat rank deficiencies in the concentration profiles and thus to allow the fitting of the pure component spectra according to Equation 4.9 is applied to three different mechanisms of increasing complexity. In the first example, the effect of dosing on the rank of **C** is presented for a second order reaction ($A + B \rightarrow C$). The dosed species leading to a rank augmentation of **C** (Strategy 3) are also discussed as a function of their initial concentrations.

As a second example, two parallel consecutive second order reactions ($A + B \rightarrow C$ and $A + C \rightarrow D$) are studied under batch conditions in order to show the impact on the rank of **C** defining a species uncoloured (Strategy 1) or providing a known pure component spectrum (Strategy 2). For the last example, three parallel consecutive reactions ($A + B \rightarrow C$, $A + C \rightarrow D$ and $A + C \rightarrow E$), all strategies (1) to (4) are analysed.

4.4.1 Example 1: a second order reaction

In this example, we apply the proposed systematic method to a second order reaction in which a species *A* reacts with a species *B* to form a third species *C*. The number of reactions is *nr* = 1 and the number of species is *ns* = 3. As only one reaction is considered, several matrices and vectors collapse to vectors and scalars respectively. For the sake of simplicity, however, the boldface capital and boldface lowercase notations for matrices and vectors respectively are maintained.

$$A + B \xrightarrow{k_1} C \quad \text{with} \quad \mathbf{k} = 0.5\ \text{L}\cdot\text{mol}^{-1}\cdot\text{min}^{-1} \tag{4.24}$$

$$\mathbf{E} = \begin{bmatrix} \overset{A}{1} & \overset{B}{1} & \overset{C}{0} \end{bmatrix},\ \mathbf{P} = \begin{bmatrix} \overset{A}{0} & \overset{B}{0} & \overset{C}{1} \end{bmatrix},\ \mathbf{N} = \mathbf{P} - \mathbf{E} = \begin{bmatrix} \overset{A}{-1} & \overset{B}{-1} & \overset{C}{1} \end{bmatrix} \tag{4.25}$$

Using Equations 4.6 to 4.8, the *ns* = 3 differential equations describing the mechanism of Equation 4.24 in concentration terms (*A*, *B*, and *C*) are replaced by *nr* = 1 differential equation expressed in terms of one extent of reaction.

$$\frac{\mathrm{d}x_{\mathrm{t},1}}{\mathrm{d}t} = k_1 \prod_{i=1}^{ns=3} c_{\mathrm{t,i}}^{e_{1,i}} - \frac{f_\mathrm{t}}{v_\mathrm{t}} x_{\mathrm{t},1} = k_1 c_{\mathrm{t},A} c_{\mathrm{t},B} - \frac{f_\mathrm{t}}{v_\mathrm{t}} x_{\mathrm{t},1} \tag{4.26}$$

Note that the term $(f_\mathrm{t}/v_\mathrm{t})x_{\mathrm{t},1}$ related to dilution in Equation 4.26 is zero under batch conditions (assuming constant density), as $f_\mathrm{t} = 0$ and $v_\mathrm{t} = \mathrm{v}_0$ for all *t*.

Differential Equation 4.26 is integrated with initial conditions $\mathbf{x_0} = \mathbf{0}$. The related vector of initial concentrations $\mathbf{c_0}$ (1 x 3), in $\text{mol}\cdot\text{L}^{-1}$, is defined as a function of the initial concentration of species *A*, denoted $c_{0,A}$ and of the stoichiometric ratio between the initial concentrations of *B* and *A*, denoted $\alpha = c_{0,B}/c_{0,A}$.

$$\mathbf{c_0} = \begin{bmatrix} \overset{A}{c_{0,A}} & \overset{B}{\alpha c_{0,A}} & \overset{C}{0} \end{bmatrix} \tag{4.27}$$

Under batch conditions, $\mathbf{C_{in}}$ is **0** (*nf* = 0) and the augmented matrix $\boldsymbol{\Omega}$ (4 x 3) is obtained by vertical stacking of $(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}$ (3 x 3) with $\mathbf{c_0}$ (1 x 3).

$$\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \left[\begin{array}{c} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c_0}} \end{array}\right] = \left[\begin{array}{ccc} \overset{A}{} & \overset{B}{} & \overset{C}{} \\ -1 & -1 & 1 \\ -1 & -1 & 1 \\ -0.5 & -0.5 & 0.5 \\ \hline c_{0,A} & \alpha c_{0,A} & 0 \end{array}\right] \tag{4.28}$$

With μ arbitrarily fixed to a value of 2, and **1** a vector of ones of dimensions (3 x 1), and **k** = 0.5 L·mol^{-1}·min^{-1}.

In the notation for $\boldsymbol{\Omega}$ used in Equation 4.28, the superscripts indicate the species that have been removed from $\boldsymbol{\Omega}$ to perform a model reduction (Strategy 1 or 2). The first superscript (before the slash) indicates which species have been defined uncoloured in Strategy 1 (∅ = no uncoloured species), whereas the second superscript (after the slash) denotes which pure spectra have been provided in Strategy 2 (∅ = no pure spectrum provided). The subscripts indicate the species that have been added to $\boldsymbol{\Omega}$ in order to augment its rank (Strategy 3 or 4). The first subscript (before the slash) indicates which species have been dosed in Strategy 3 (∅ = batch conditions) and the second subscript (after the slash) denotes the species for which initial concentrations have been varied in Strategy 4 (∅ = no additional experiment).

As the second row of $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (4 x 3) is identical to the first one and the third one is a multiple of the first one, the rank of $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ and thus of **C** is two (*nc* = 2). This is in agreement with the prediction of Amrhein et al (Equation 4.14). Figure 4.5a presents the rank deficient concentration profiles of the three species involved in the mechanism of Equation 4.24 (**k** = 0.5 L·mol^{-1}·min^{-1}) for initial conditions $c_{0,A}$ = 0.5 molL^{-1} and $c_{0,B}$ = 1 mol·L^{-1} (α = 2).

According to the rank-nullity theorem (Equation 4.12) and the property of pseudo-equivalence (Equation 4.13), the dimension of the kernel of $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ and of **C** is *ns*–*nc*=3–2=1. The kernel of $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ is computed by SVD and results in an orthonormal basis of dimension 1 that depends on the stoichiometric ratio α.

$$\ker \mathbf{C} = \ker \mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} \overset{A}{-\alpha} & \overset{B}{1} & \overset{C}{1-\alpha} \end{bmatrix}^{\mathrm{T}} \tag{4.29}$$

Equation 4.29 can be verified with Figure 4.5a (α = 2), as all three concentration profiles at any time *t* are related by Equation 4.30, which is deduced from the definition of the kernel (see Section 4.3.1.1).

$$\mathbf{C}(\ker \mathbf{C}) = \mathbf{0} \;\Rightarrow\; -\alpha c_{\mathrm{t},A} + 1c_{\mathrm{t},B} + (1-\alpha)c_{\mathrm{t},C} = -2c_{\mathrm{t},A} + c_{\mathrm{t},B} - c_{\mathrm{t},C} = 0 \tag{4.30}$$

Based on $\ker \mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (3 x 1) (Equation 4.29) and on the value of the stoichiometric ratio α, the species to be dosed in order to reach a full spectral resolution can be identified.

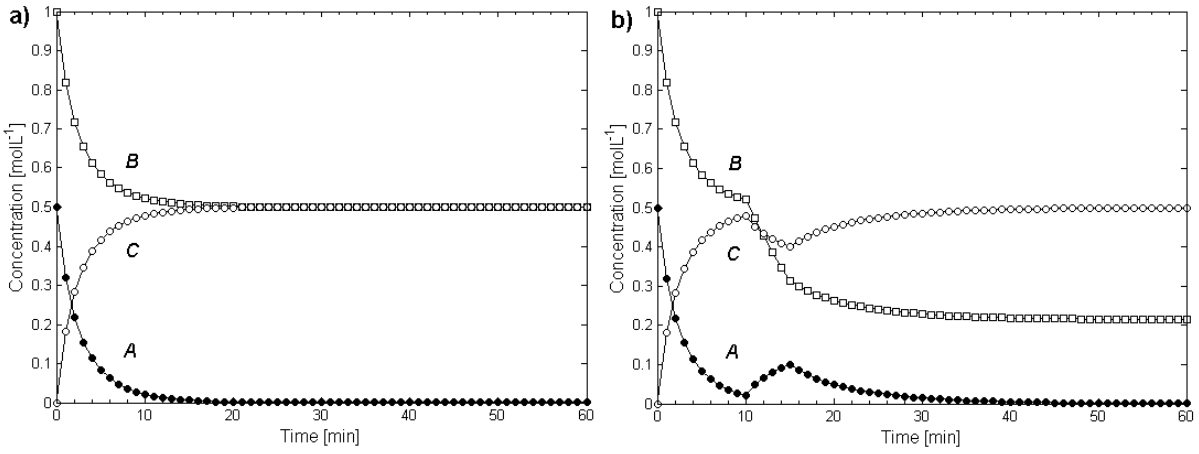

Figure 4.5: Time resolved concentration profiles for the mechanism described in Equation 4.24 (**k** = 0.5 L·mol^{-1}·min^{-1}) under batch (a) and semi-batch (b) conditions. In (a) and (b), species *A* (●) and *B* (□) are initially mixed with $c_{0,A}$ = 0.5 mol·L^{-1}, $c_{0,B}$ = 1 mol·L^{-1} (α = 2) and v_0 = 25·10^{-3} L, and lead to the product *C* (○). In (b) species *A* is subsequently dosed at *t* = 10 min during 5 min with a dosing rate of 2·10^{-3} L·min^{-1} and a dosing concentration $c_{in,A}$ = 0.5 mol·L^{-1}.

4.4.1.1 Rank augmentation by dosing (Strategy 3)

The element in the kernel corresponding to species *C* is zero only when α = 1, i.e. under stoichiometric conditions (see Equation 4.29). This indicates that under these conditions, only species *A* or *B* can be dosed in order to break the rank deficiency in matrix **C**. Under non-stoichiometric conditions (α = 1), however, any of the three species could be dosed in order to augment the rank of $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (or **C**).

For the particular example of Figure 4.5a, if species *A* is dosed, matrix $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ is augmented by an additional row containing only the concentration of the dosed species ($c_{in,A}$). The rank of the resulting matrix, denoted $\boldsymbol{\Omega}\big|_{A/\varnothing}^{\varnothing/\varnothing}$ (5 x 3), is augmented to full rank (see Equation 4.31) and its kernel only contains the trivial solution (**0**).

Figure 4.5b presents the linearly independent concentration profiles of the three species involved in the mechanism of Equation 4.24 for the same initial conditions as in Figure 4.5a but under semi-batch conditions (species *A* is dosed with a concentration $c_{in,A}$ = 0.5 mol·L^{-1}).

$$\boldsymbol{\Omega}\big|_{A/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^{\mathrm{T}}}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c_0}} \\ \hline \mathbf{C_{in}} \end{bmatrix} = \left[\begin{array}{ccc} A & B & C \\ -1 & -1 & 1 \\ -1 & -1 & 1 \\ -0.5 & -0.5 & 0.5 \\ \hline c_{0,A} & \alpha c_{0,A} & 0 \\ \hline c_{in,A} & 0 & 0 \end{array}\right] \;;\; \ker \boldsymbol{\Omega}\big|_{A/\varnothing}^{\varnothing/\varnothing} = \ker \mathbf{C} = \mathbf{0} \qquad (4.31)$$

4.4.2 Example 2: a mechanism with two parallel consecutive reactions

The mechanism presented in Example 1 ($A + B \rightarrow C$) is coupled to a second parallel and consecutive reaction in which species *A* reacts with *C* to form a product *D* (see Figure 4.1). The number of species is *ns* = 4 (*A*, *B*, *C* and *D*) and the number of reactions is *nr* = 2. Vector **k** is comprised of two rate constants that are to be determined by non-linear optimisation. The mechanism of Figure 4.1 can be translated in *nr* = 2 differential equations. Thus, differential Equation 4.26 of Example 1, without the term $(f_t/v_t)x_{t,1}$ as only batch conditions are considered here, is coupled to a second differential equation.

$$\frac{\mathrm{d}x_{\mathrm{t},2}}{\mathrm{d}t} = k_2 \prod_{i=1}^{ns=4} c_{\mathrm{t,i}}^{e_{2,\mathrm{i}}} = k_2 c_{\mathrm{t},A} c_{\mathrm{t},C} \tag{4.32}$$

Differential Equations 4.26 and 4.32 are integrated simultaneously with initial conditions $\mathbf{x_0} = \mathbf{0}$. As before, the related vector of initial concentrations $\mathbf{c_0}$ (1 x 4), in mol·L^{-1}, is defined as a function of the initial concentration of *A*.

$$\mathbf{c_0} = \begin{bmatrix} \overset{A}{c_{0,A}} & \overset{B}{\alpha c_{0,A}} & \overset{C}{0} & \overset{D}{0} \end{bmatrix} \tag{4.33}$$

The augmented matrix $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 4) is obtained by vertical stacking of $(\mu\mathbf{1})^{\bullet\mathbf{E}^{\mathrm{T}}}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}$ (4 x 4) with $\mathbf{c_0}$ (1 x 4). Note that, as the rate constants are not a priori known, Equation 4.17 is used with a symbolic vector **k** of dimensions (1 x 2), i.e. $\mathbf{k} = [k_1, k_2]$.

$$\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^{\mathrm{T}}}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c_0}} \end{bmatrix} = \left[\begin{array}{cccc} \overset{A}{} & \overset{B}{} & \overset{C}{} & \overset{D}{} \\ -\mu k_1 - \mu k_2 & -\mu k_1 & \mu k_1 - \mu k_2 & \mu k_2 \\ -\mu k_1 - k_2 & -\mu k_1 & \mu k_1 - k_2 & k_2 \\ -k_1 - \mu k_2 & -k_1 & k_1 - \mu k_2 & \mu k_2 \\ -k_1 - k_2 & -k_1 & k_1 - k_2 & k_2 \\ \hline c_{0,A} & \alpha c_{0,A} & 0 & 0 \end{array}\right] \tag{4.34}$$

With μ any positive scalar different from 0 or 1, and **1** a matrix of ones of dimensions (4 x 2).

The rank of **C**, as predicted by Equation 4.14 and calculated from $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, is *nc* = 3, i.e. at most three species are allowed to be set to coloured in order to circumvent rank deficiency. Using the rank-nullity theorem (Equation 4.12), the dimension of the kernel of **C** is (*ns*–*nc*) = (4–3) = 1. Accordingly, the orthonormal basis for the kernel of $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, computed by SVD, results in a vector space of dimension 1 which depends on α.

$$\ker\ \boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} \overset{A}{-\alpha} & \overset{B}{1} & \overset{C}{1-\alpha} & \overset{D}{1-2\alpha} \end{bmatrix}^{\mathrm{T}} \tag{4.35}$$

Note that for this example the kernel of $\boldsymbol{\Omega}$ neither depends on μ nor on **k**.

The few Matlab lines that are required for the setup of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ and the calculation of its kernel (Equation 4.35) are provided in Section 4.6.2 (Appendix) for this example. The kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ can be used to determine which species to define uncoloured (Section 4.3.2.1) or which pure component spectrum to provide in order to reach full spectral resolution (Section 4.3.2.2). As $\ker \mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (4 x 1) is a function of the stoichiometric ratio α, the species to include in Strategies (1) and (2) will depend on the ratio of the initial concentrations.

4.4.2.1 Model reduction by defining uncoloured species (Strategy 1)

As shown by the rank of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, *nc* = 3 appropriate coloured species have to be chosen in order to compute the pseudo-inverse of $\mathbf{C_c}$ and so to optimise the two unknown rate constants. The (*ns*–*nc*) = *nu* = 1 uncoloured species can be chosen amongst the non-zero rows of $\ker \mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (see Scheme 4.1a and Equation 4.35), i.e. amongst the species *A*, *B* or *D* under stoichiometric conditions (α = 1), amongst *A*, *B* or *C* under two fold excess of A (α = 1/2) or amongst any of the four species outside from these two particular conditions. For example, if species *A* is set uncoloured, the column corresponding to *A* is removed from the augmented matrix $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, leading for any value of α to a matrix $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{A/\varnothing}$ that is of full rank and an associated kernel only comprised of the trivial solution (**0**).

In order to predict the impact of defining *A* uncoloured on the linear combinations of the true pure component spectra (inherent problem in the fitted component spectra when Strategy (1) is used), the linear dependencies in the rows of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ are removed. For this, iterative Scheme 4.1b is employed and all five rows are first considered linearly independent (initialisation step). The first linearly dependent row is identified amongst the non-zero rows of $\ker \mathbf{\Omega}^{\mathrm{T}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 2), as shown in the first matrix from the left of Equation 4.36.

$$\ker\ \mathbf{\Omega}^{\mathrm{T}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -\mu & -(\mu+1) \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{n/a} \\ 1 \\ 1 \\ -(\mu+1) \\ 0 \end{bmatrix} \Rightarrow \mathbf{0} \tag{4.36}$$

According to $\ker\ \mathbf{\Omega}^{\mathrm{T}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (first matrix (5 x 2) in Equation 4.36), any row except the last one (corresponding to $\mathbf{c_0}$) can be removed. If, for example, the first row is removed, the kernel computed from this augmented matrix is shown in the second matrix (4 x 1) of Equation 4.36. Note that n/a denotes the entry corresponding to the first row that does not exist anymore in the kernel.

Subsequently, based on the zero entries of the second matrix (i.e. vector) in Equation 4.36, any row except the last one can be removed. If, for example, the second row is removed, the kernel of the augmented matrix is only comprised by the trivial solution (**0**). Removing the first and the second row in $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 4) leads to a matrix denoted $\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ of dimensions (3 x 4) that can be used in Equation 4.20 in order to predict the coefficients for the linear combinations of the true pure component spectra. These linear combinations are calculated, based on the fact that species *A* was set uncoloured, i.e. eliminating the column corresponding to species *A* in $\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, leading to $\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{A/\varnothing}$.

$$\mathbf{\Delta} = \left(\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{A/\varnothing}\right)^{-1} \mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{array}{c} \begin{matrix} \quad A & \quad B & \quad C & \quad D \end{matrix} \\ \begin{bmatrix} \alpha^{-1} & 1 & 0 & 0 \\ \alpha^{-1}-1 & 0 & 1 & 0 \\ \alpha^{-1}-2 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} \\ \text{'}B\text{'} \\ \text{'}C\text{'} \\ \text{'}D\text{'} \end{matrix} \tag{4.37}$$

With the rows '*B*', '*C*' and '*D*' indicating the coloured species for which fitted component spectra are obtained, and the columns *A*, *B*, *C*, *D* denoting the true absorbing species for which true pure component spectra exist. Note that, for this example, $\mathbf{\Delta}$ neither depends on μ nor on **k**.

Equation 4.37 indicates the linear combinations of the true pure component spectra **A** that lead to the fitted pure component spectra $\mathbf{A_c}$.

To illustrate Equation 4.37, four true pure component spectra **A** were generated using Gaussian functions as shown in Figure 4.6a. These were used to simulate a kinetic data matrix **Y**, using a set of concentration profiles (α = 2) corresponding to the mechanism given in Figure 4.1.

Figure 4.6b shows the linear combinations inherent to the fitted component spectra of the coloured species after optimisation (including Equation 4.9) and Equation 4.37 can be interpreted as follows: the fitted component spectrum of the coloured species '*B*' is a linear combination of 0.5 x true pure component spectrum of *A* with 1 x true pure component spectrum of *B*; the fitted spectrum of the coloured species '*C*' is a linear combination of -0.5 x true *A* with 1 x true *C*; and the fitted spectrum of the coloured species '*D*' is a linear combination of -1.5 x true *A* with 1 x true *D*.

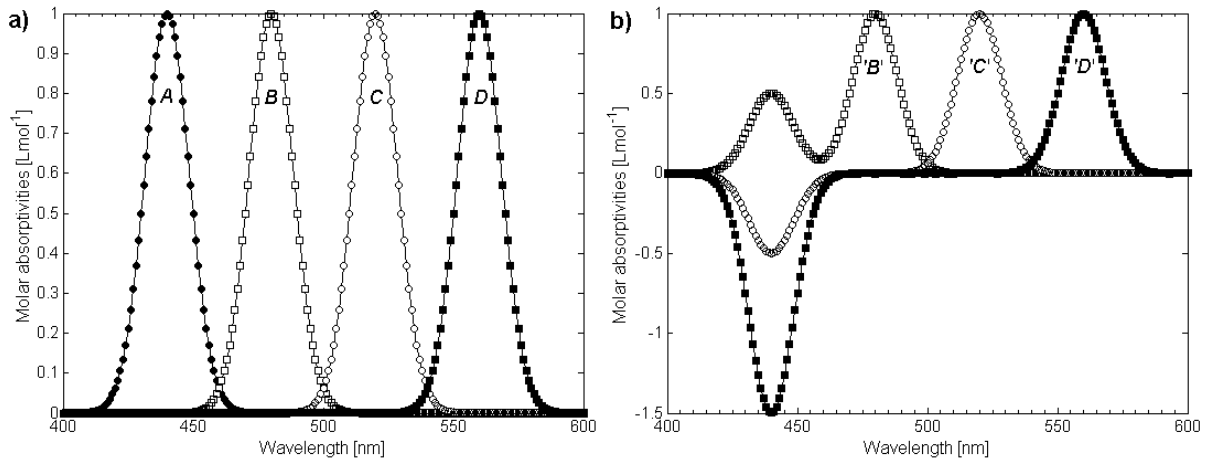

Figure 4.6: Simulated true pure (a) and fitted (b) component spectra for the mechanism described in Figure 4.1 with $\alpha = 2$. The *ns* = 4 simulated pure component spectra (a) were generated using Gaussian functions centred at 440 nm (species *A* ●), 480 nm (*B* □), 520 nm (*C* ○), 560 nm (*D* ◘) with a constant half width of 20 nm and a maximum intensity of 1. The fitted component spectra (b) of '*B*', '*C*' and '*D*' were obtained by defining species *A* uncoloured and applying Equation 4.9.

4.4.2.2 Model reduction by providing a pure spectrum (Strategy 2)

When α = 1, i.e. under stoichiometric conditions, the element in the kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ corresponding to species *C* becomes zero (see Equation 4.35) and thus either the pure component spectrum of species *A*, *B* or *D* can be provided in order to obtain full resolution of all component spectra. However, when α = 1/2 (a two fold excess of *A*), the entry for *D* in the kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ becomes zero and only incorporating the pure component spectrum of *A*, *B* or *C* leads to a full spectral resolution. Apart from these two particular conditions (α = 1 or α = 1/2), the pure component spectrum of any of the four species can be provided. For example, if the spectrum of *B* is provided, the column corresponding to *B* is removed.

$$\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/B} = \left[\begin{array}{ccc} \overset{A}{} & \overset{C}{} & \overset{D}{} \\ -\mu k_1 - \mu k_2 & \mu k_1 - \mu k_2 & \mu k_2 \\ -\mu k_1 - k_2 & \mu k_1 - k_2 & k_2 \\ -k_1 - \mu k_2 & k_1 - \mu k_2 & \mu k_2 \\ -k_1 - k_2 & k_1 - k_2 & k_2 \\ \hline c_{0,A} & 0 & 0 \end{array}\right] \tag{4.38}$$

The kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/B}$ only contains the trivial solution (**0**) and thus, all fitted component spectra are resolved. In other words, matrix $\mathbf{\Delta}$ for the linear combinations of the true pure component spectra equals the identity matrix **I**.

4.4.3 Example 3: a mechanism with three parallel consecutive reactions

Finally, a complex kinetic scheme is studied in a systematic way and different experimental conditions are discussed in order to reach full spectral resolution. The mechanism presented in Example 2 is coupled to a third parallel reaction in which species *A* reacts with *C* to form a product *E*, as shown in Equation 4.39. This kinetic scheme involves *nr* = 3 reactions and *ns* = 5 species (*A*, *B*, *C*, *D* and *E*), and results in matrices **E**, **P** and **N**, all of dimensions (3 x 5), which are defined in Equation 4.40. The three associated rate constants are considered as unknown and may be determined by non-linear optimisation.

$$A + B \xrightarrow{k_1} C \;,\; A + C \xrightarrow{k_2} D \;,\; A + C \xrightarrow{k_3} E \tag{4.39}$$

$$\mathbf{E} = \begin{matrix} A & B & C & D & E \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{P} = \begin{matrix} A & B & C & D & E \end{matrix} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{N} = \begin{matrix} A & B & C & D & E \end{matrix} \begin{bmatrix} -1 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{bmatrix} \tag{4.40}$$

The mechanism of Equation 4.39 can be written in *nr* = 3 differential equations. Thus, differential Equation 4.26 (Example 1) and differential Equation 4.32 (Example 2) are coupled to a third differential equation.

$$\frac{\mathrm{d}x_{\mathrm{t},3}}{\mathrm{d}t} = k_3 \prod_{i=1}^{ns=5} c_{\mathrm{t,i}}^{e_{3,i}} - \frac{f_{\mathrm{t}}}{v_{\mathrm{t}}} x_{\mathrm{t},3} = k_3 c_{\mathrm{t},A} c_{\mathrm{t},C} - \frac{f_{\mathrm{t}}}{v_{\mathrm{t}}} x_{\mathrm{t},3} \tag{4.41}$$

As batch and semi-batch (i.e. Strategy 3) conditions are both considered in this third example, differential Equations 4.26, 4.32 and 4.41 are defined including a possible dosing. If batch conditions are desired, the flow rate f_t is set to zero for all times ($v_t = v_0$) and in such case the terms $(f_t/v_t)x_{t,1}$, $(f_t/v_t)x_{t,2}$ and $(f_t/v_t)x_{t,3}$ vanish.

Differential Equations 4.26, 4.32 and 4.41 are integrated simultaneously with initial conditions $\mathbf{x_0} = \mathbf{0}$. The related vector of initial concentrations $\mathbf{c_0}$ (1 x 5), in mol·L^{-1}, is defined as a function of the initial concentration of *A*.

$$\mathbf{c_0} = \begin{matrix} A & B & C & D & E \end{matrix} \begin{bmatrix} c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \end{bmatrix} \tag{4.42}$$

As the vector of rate constants is unknown, Equation 4.16 is used with a symbolic vector **k** of dimensions (1 x 3), leading to an augmented matrix of dimensions (6 x 5), obtained by stacking column-wise $(\mu \mathbf{1})^{\bullet \mathbf{E}^{\mathrm{T}}}$ DIAG$(\mathbf{k})\mathbf{N}$ (5 x 5) and $\mathbf{c_0}$ (1 x 5).

$$\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \left[\frac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c_0}}\right] = \left[\begin{array}{ccccc} A & B & C & D & E \\ -2k_1-2k_2-2k_3 & -2k_1 & 2k_1-2k_2-2k_3 & 2k_2 & 2k_3 \\ -2k_1-k_2-k_3 & -2k_1 & 2k_1-k_2-k_3 & k_2 & k_3 \\ -k_1-2k_2-2k_3 & -k_1 & k_1-2k_2-2k_3 & 2k_2 & 2k_3 \\ -k_1-k_2-k_3 & -k_1 & -k_1-k_2-k_3 & k_2 & k_3 \\ -k_1-k_2-k_3 & -k_1 & -k_1-k_2-k_3 & k_2 & k_3 \\ \hline c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \end{array}\right] \quad (4.43)$$

With **1** a matrix of ones of dimensions (5 x 3), **k** = [k_1, k_2, k_3] and μ arbitrarily fixed to a value of 2.

As presented in Equation 4.43 and in Table 4.2, $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ and thus **C** (*nt* x *ns*) are rank deficient under batch conditions, i.e. (*nc* = 3) < (*ns* = 5), and the dimension of their kernels is (*ns*–*nc*) = (5–3) = 2. An orthonormal basis of dimension 2 for ker $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 2) (and ker **C**) is computed by SVD depending on α and **k**, as presented in Table 4.3.

For this example, the agreement between the kernel of $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ and kernel of **C** is also presented in Section 4.6.3 (Appendix). Due to the rank deficiency in $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (and in **C**), $\mathbf{C}^+$ in Equation 4.9 cannot be computed and the *ns* pure component spectra cannot be calculated.

In the following, some options, summarised in Table 4.2 and Table 4.3, are presented in order to circumvent this mathematical drawback, e.g. by applying each strategy (1, 2, 3 or 4) individually or combining Strategy (1) with strategies (2, 3 or 4). Note that many more permutations are possible (strategies 2 & 3, 2 & 4, or 3 & 4) and other species can be chosen within the individual strategies. The best strategy somewhat depends on the experimental constraints, such as accessibility of known spectra or feasibility of dosing a certain species. Importantly, their outcome can all be predicted (see Section 4.6.4 – Appendix).

Table 4.2: Dimension of the image and of the kernel of $\boldsymbol{\Omega}$ (mechanism of Equation 4.39) when strategies (1), (2), (3) and (4) are applied.

	$\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/\varnothing}$ [1]	$\boldsymbol{\Omega}\big\|^{B/\varnothing}_{\varnothing/\varnothing}$	$\boldsymbol{\Omega}\big\|^{B,E/\varnothing}_{\varnothing/\varnothing}$	$\boldsymbol{\Omega}\big\|^{\varnothing/A}_{\varnothing/\varnothing}$	$\boldsymbol{\Omega}\big\|^{D/A}_{\varnothing/\varnothing}$	$\boldsymbol{\Omega}\big\|^{\varnothing/A,D}_{\varnothing/\varnothing}$	$\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{A/\varnothing}$	$\boldsymbol{\Omega}\big\|^{D/\varnothing}_{A/\varnothing}$	$\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{A,D/\varnothing}$	$\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/B}$	$\boldsymbol{\Omega}\big\|^{E/\varnothing}_{\varnothing/B}$	$\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/B,E}$
Strategy	None	(1)	(1), (1)	(2)	(1), (2)	(2), (2)	(3)	(1), (3)	(3), (3)	(4)	(1), (4)	(4), (4)
nc = rank($\boldsymbol{\Omega}$)	3	3	3	3	3	3	4	4	5	4	4	5
dim(ker $\boldsymbol{\Omega}$)	2	1	0	1	0	0	1	0	0	1	0	0
Number of species	5	4	3	4	3	3	5	4	5	5	4	5
Resolved pure spectra	n/a[2]	n/a	C[3]	n/a	A, B, C	all	n/a	A, B, C	all	n/a	A, B, C	all

[1] $\boldsymbol{\Omega}\big|^{\text{uncoloured species in strategy (1) / pure spectra provided in strategy (2)}}_{\text{dosed species in strategy (3) / initial concentrations varied in strategy (4)}}$

[2] n/a indicates that the pseudo-inverse of the concentration matrix cannot be computed and thus the linear regressing step to calculate the component spectra **A** cannot be performed.

[3] The pure spectrum of species C is resolved when $\alpha = 1$.

Table 4.3: Orthonormal basis for the kernel of $\boldsymbol{\Omega}$ (mechanism of Equation 4.39) when strategies (1), (2), (3) and (4) are applied.

Species	ker $\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/\varnothing}$ [1], [2]		ker $\boldsymbol{\Omega}\big\|^{B/\varnothing}_{\varnothing/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{B,E/\varnothing}_{\varnothing/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/A}_{\varnothing/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{D/A}_{\varnothing/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/A,D}_{\varnothing/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{A/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{D/\varnothing}_{A/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{A,D/\varnothing}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/B}$	ker $\boldsymbol{\Omega}\big\|^{E/\varnothing}_{\varnothing/B}$	ker $\boldsymbol{\Omega}\big\|^{\varnothing/\varnothing}_{\varnothing/B,E}$
A	$-\alpha$	0	0	0	n/a[4]	n/a[4]	n/a[4]	0	0	0	0	0	0
B	1	0	n/a[3]	n/a[3]	0	0	0	0	0	0	0	0	0
C	$1-\alpha$	0	0	0	0	0	0	0	0	0	0	0	0
D	$(1+k')(1-2\alpha)$	$-k'$	$-k'$	0	$-k'$	n/a[3]	n/a[4]	$-k'$	n/a[3]	0	$-k'$	0	0
E	0	1	1	n/a[3]	1	0	0	1	0	0	1	n/a[3]	0

[1] $\boldsymbol{\Omega}\big|^{\text{uncoloured species in strategy (1) / pure spectra provided in strategy (2)}}_{\text{dosed species in strategy (3) / initial concentrations varied in strategy (4)}}$

[2] $k' = k_3/k_2$

[3] n/a indicates that the species was excluded from $\boldsymbol{\Omega}$, as it was set uncoloured.

[4] n/a indicates that the species was excluded from $\boldsymbol{\Omega}$, as its pure spectrum was provided.

4.4.3.1 Model reduction by defining uncoloured species (Strategy 1)

According to the method described in Scheme 4.1a, first all five species are considered coloured. All rows of ker $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 2) contain non-zero entries for any value of α (see Table 4.3) except for species *C* which contain a row full of zeros when α = 1. Thus, the first uncoloured species can be freely chosen amongst the five species if $\alpha \neq 1$ or amongst *A*, *B*, *D* and *E* if α = 1. If e.g. species B is set uncoloured, the column associated to species *B* is removed from $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 5), leading to a matrix $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{B/\varnothing}$ (6 x 4). Thus the number of coloured species becomes (*ns*–*nu*) = (5–1) = 4. The dimension of ker $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{B/\varnothing}$ (4 x 1) is only one, as shown in Table 4.2. The impact of defining another uncoloured species (*A*, *C*, *D* or *E*) on the kernel of $\mathbf{\Omega}$ is reported in Section 4.6.4 (Appendix).

A second appropriate uncoloured species, deduced from the non-zero entries of ker $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{B/\varnothing}$ (see Table 4.3), can be chosen amongst species *D* or *E*. If e.g. species *E* is set uncoloured, the column corresponding to species *E* is removed from $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{B/\varnothing}$ (6 x 4) leading to a matrix $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{B,E/\varnothing}$ (6 x 3), whose kernel only contains **0**. Thus, an appropriate set of (*ns*–*nu*) = (5–2) = 3 coloured species is '*A*', '*C*' and '*D*'.

In order to predict the impact on the fitted component spectra of defining species *B* and *E* uncoloured, three rows need to be removed sequentially from $\mathbf{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 5) according to iterative Scheme 4.1b using $\mathbf{\Omega}^{\mathrm{T}}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (5 x 6). The first row to be removed is identified amongst the non-zero rows of ker $\mathbf{\Omega}^{\mathrm{T}}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 3) (first matrix from the left in Equation 4.44), i.e. from row one to five.

$$\ker\ \boldsymbol{\Omega}^{\mathrm{T}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} -3 & -1 & -1 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{n/a} & \text{n/a} \\ -1 & -1 \\ -1 & -1 \\ 0 & 3 \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{n/a} \\ -1 \\ -1 \\ \text{n/a} \\ 3 \\ 0 \end{bmatrix} \Rightarrow \mathbf{0} \tag{4.44}$$

Removing for example row one and re-computing the kernel leads to the second matrix (5 x 2) in Equation 4.44. The second row to be removed can be chosen amongst rows two to five of $\boldsymbol{\Omega}$. If, for example, row four is removed and the kernel is re-calculated, this leads to the third matrix (4 x 1) in Equation 4.44. The third row to be removed can finally be chosen amongst rows two, three and five. Removing for example row five of $\boldsymbol{\Omega}$ leads to a kernel only comprised of **0**. Thus, rows two, three and six form an appropriate set of linearly independent rows that can be used to reduce $\boldsymbol{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 5) to $\boldsymbol{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (3 x 5), and to calculate the underlying linear combinations of the true pure component spectra that form the fitted component spectra.

$$\boldsymbol{\Delta} = \left(\boldsymbol{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{B,E/\varnothing}\right)^{-1} \boldsymbol{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{matrix} \\ \\ \\ \end{matrix} \begin{matrix} A & B & C & D & E \\ \end{matrix} \begin{bmatrix} 1 & \alpha & 0 & 0 & 0 \\ 0 & \alpha-1 & 1 & 0 & 0 \\ 0 & (1+k')(2\alpha-1) & 0 & 1 & k' \end{bmatrix} \begin{matrix} \text{'A'} \\ \text{'C'} \\ \text{'D'} \end{matrix} \tag{4.45}$$

With $k' = k_3/k_2$.

Note that, in contrast to Example 2, $\boldsymbol{\Delta}$ now depends on **k**. As shown in Equation 4.45, whatever the value of α, none of the fitted component spectra will be resolved except the one of species *C* when $\alpha = 1$. As spectral interpretation is very important in order to validate a reaction mechanism, particularly if it is a complex one, spectral resolution on the species level is highly desired. In the next sections, we will show how this can be achieved by an experimental design following our systematic approach.

4.4.3.2 Model reduction by providing known pure spectra (Strategy 2)

According to the entries in $\ker \left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (see Table 4.3), any pure component spectrum can be provided in Strategy (2), except the pure spectrum of *C* when $\alpha = 1$. Assume for example the pure spectrum of reactant *A* is provided. In this case, the column corresponding to species *A* is removed from $\left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 5), resulting in a matrix $\left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ (6 x 4) and in a reduced number of species, i.e. (*ns*–*nks*) = (5–1) = 4. The rank of $\left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ is *nc* = 3 and the dimension of $\ker \left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ (6 x 1) is 1, as shown in Table 4.2. The impact on the kernel of $\mathbf{\Omega}$ of providing any other pure component spectrum (*B*, *C*, *D*, or *E*) is reported in Section 4.6.4 (Appendix). An orthonormal basis for the kernel of $\left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ is calculated in Table 4.3.

As $\dim\left(\ker \left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}\right)$ is not zero and the rank of $\left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ (*nc* = 3) is lower than the number of species (*ns*–*nks* = 4), rank deficiency in the concentration matrix is not yet removed. In order to do so, either another pure component spectrum can be provided, i.e. continue with Strategy (2) or, if none is accessible, an appropriate species can be defined uncoloured, i.e. combining strategies (1) and (2).

If, for example, Strategy (2) is applied a second time, the species to supply a pure component spectrum for can be selected amongst the species *D* and *E* in order to reach full spectral resolution.

If, however, Strategy (1) is to be followed, Beer's law (Equation 4.1) has to be reduced by defining *nc* = 3 coloured species amongst the (*ns*–*nks*) = 4 species *B*, *C*, *D* and *E*. Note however that species *B* and *C* cannot be set uncoloured as their corresponding entries in $\ker \left.\mathbf{\Omega}\right|_{\varnothing/\varnothing}^{\varnothing/A}$ are zero. For example, species '*B*', '*C*' and '*E*' form an appropriate set of coloured species, and species *D* can be defined uncoloured. Another possible set of coloured species is '*B*', '*C*' and '*D*', with species *E* uncoloured.

The linear combinations of the true pure component spectra corresponding to the coloured species '*B*', '*C*' and '*E*' can be predicted using Equation 4.20 and independent rows of the augmented matrix (rows two, three and six as discussed in section 4.4.3.1). Recall, as the pure component spectrum of *A* was initially provided, the linear combinations predicted in Equation 4.46 do not contain any column related to *A*.

$$\mathbf{\Delta} = \left(\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{D/A}\right)^{-1}\mathbf{\Omega}_{\mathbf{ind}}\Big|_{\varnothing/\varnothing}^{\varnothing/A} = \begin{array}{c} \begin{matrix} B & C & D & E \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \left(k'\right)^{-1} & 1 \end{bmatrix} \end{array} \begin{matrix} \\ 'B' \\ 'C' \\ 'E' \end{matrix} \tag{4.46}$$

With $k' = k_3/k_2$.

Note that the fitted pure component spectra of *B* and *C* are always resolved, as '*B*' is only a linear combination with itself (*B*) and so does '*C*' with itself (*C*). In addition, the fitted spectrum of the coloured species '*E*' is always a linear combination between the true spectrum of *D* and it-self (*E*).

4.4.3.3 Rank augmentation by dosing (Strategy 3)

Based on the method described in Scheme 4.3 and on the kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (see Table 4.3), any species could potentially be dosed in order to augment the rank of matrix **C**, except species *C* when α = 1. Assuming, for example, reactant *A* is dosed (with a dosing concentration $c_{in,A}$), this leads to an augmented matrix $\mathbf{\Omega}\Big|_{A/\varnothing}^{\varnothing/\varnothing}$ of dimensions (7 x 5). The impact on the kernel of $\mathbf{\Omega}$ of dosing any other species (*B*, *C*, *D* or *E*) is reported in Section 4.6.4 (Appendix).

$$\mathbf{\Omega}\Big|_{A/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} \dfrac{\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^{\mathrm{T}}}\mathrm{DIAG}\left(\mathbf{k}\right)\mathbf{N}}{\mathbf{c_0}} \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}}{\mathbf{C_{in}}} \end{bmatrix} = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \left[\begin{array}{c} \mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} \\ \hline c_{in,A} \quad 0 \quad 0 \quad 0 \quad 0 \end{array}\right] \end{array} \tag{4.47}$$

The rank of $\mathbf{\Omega}\Big|_{A/\varnothing}^{\varnothing/\varnothing}$ (7 x 5) is *nc* = 4 and the dimension of its kernel is 1, as shown in Table 4.2. An orthonormal basis for $\ker\ \mathbf{\Omega}\Big|_{A/\varnothing}^{\varnothing/\varnothing}$ (5 x 1) is calculated by SVD and given in Table 4.3.

As *A* has been dosed, this species is not linearly dependent on any other species and its entry is zero in the kernel. Due to the parallel reaction with *B* forming product *C*, dosing *A* also breaks the linear dependency between *B* and *C*.

As $\dim\left(\ker \boldsymbol{\Omega}\big|_{A/\varnothing}^{\varnothing/\varnothing}\right)$ is not zero and the rank of $\boldsymbol{\Omega}\big|_{A/\varnothing}^{\varnothing/\varnothing}$ (*nc* = 4) is lower than the number of species (*ns* = 5), the fitted component spectra are not fully resolved. Two options (among others) to allow computing the pseudo-inverse of **C** could be to dose another species (Strategy 3 again) or to define one uncoloured species (Strategy 1).

When Strategy (3) is applied again, in order to reach full spectral resolution, the next species to be dosed has to be chosen amongst *D* or *E*. When Strategy (1) is applied in combination with Strategy (3), Beer's law (Equation 4.1) is reduced by defining *nc* = 4 coloured species amongst the *ns* = 5 species *A*, *B*, *C*, *D* and *E*. Due to the previous dosing of *A*, species *A*, *B* and *C* cannot be set uncoloured. A possible set of coloured species is thus '*A*', '*B*', '*C*' and '*E*', species *D* being set uncoloured. Another possible set of coloured species could be '*A*', '*B*', '*C*' and '*D*' (i.e. species *E* uncoloured). If species *D* is defined uncoloured, the linear combinations of the true pure component spectra can be predicted using Equation 4.20 and independent rows of the augmented matrix (in addition to the independent rows two, three and six discussed in section 4.4.3.1, row seven is also independent from the others as it corresponds to the dosing of species *A*).

$$\boldsymbol{\Delta} = \left(\boldsymbol{\Omega}_{\mathbf{ind}}\big|_{A/\varnothing}^{D/\varnothing}\right)^{-1} \boldsymbol{\Omega}_{\mathbf{ind}}\big|_{A/\varnothing}^{\varnothing/\varnothing} = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \left[\begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \left(k'\right)^{-1} & 1 \end{matrix}\right] \end{array} \begin{matrix} \\ 'A' \\ 'B' \\ 'C' \\ 'E' \end{matrix} \tag{4.48}$$

With $k' = k_3/k_2$.

As predicted, the fitted component spectra of species *A*, *B* and *C* are resolved whereas the component spectrum of the coloured species '*E*' has a contribution due to species *D*, the one that was set to uncoloured.

4.4.3.4 Rank augmentation by varying initial concentrations (Strategy 4)

Based on the method described in Scheme 4.4 and on the kernel of $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (see Table 4.3), in order to augment the rank of **C** the additional *ne* = 1 experiment can be performed by varying the initial concentration of either species except the concentration of species *C* when α = 1. If the ratio of the initial concentrations of *B* and *A* is varied from α to β, one additional row of initial conditions is added to $\boldsymbol{\Omega}\big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$ (6 x 5) leading to a matrix $\boldsymbol{\Omega}\big|_{\varnothing/B}^{\varnothing/\varnothing}$ of dimensions (7 x 5). The impact on the kernel of $\boldsymbol{\Omega}$ of varying other initial concentrations (for *A*, *C*, *D* or *E*) is reported in Section 4.6.4 (Appendix).

$$\boldsymbol{\Omega}\big|_{\varnothing/B}^{\varnothing/\varnothing} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c_0^{exp_1}}} \\ \hline \mathbf{c_0^{exp_2}} \end{bmatrix} = \begin{bmatrix} \overset{A}{} & \overset{B}{} & \overset{C}{} & \overset{D}{} & \overset{E}{} \\ \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\,\mathrm{DIAG}(\mathbf{k})\mathbf{N}}{} \\ \hline c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \\ c_{0,A} & \beta c_{0,A} & 0 & 0 & 0 \end{bmatrix} \tag{4.49}$$

Where $\mathbf{c_0^{exp_1}}$ and $\mathbf{c_0^{exp_2}}$ denote the vectors of initial concentrations for the first and the second experiment respectively.

The rank of $\boldsymbol{\Omega}\big|_{\varnothing/B}^{\varnothing/\varnothing}$ is 4 and the dimension of its kernel is 1, as shown in Table 4.2. An orthonormal basis for $\ker \boldsymbol{\Omega}\big|_{\varnothing/B}^{\varnothing/\varnothing}$ (5 x 1) is calculated by SVD in Table 4.3. As the stoichiometric ratio between species *A* and *B* was varied, their corresponding stacked concentration profiles are now linearly independent, i.e. their respective entries are zero in the kernel. Note that the kernel does not depend on the stoichiometric ratios α or β and that the linear dependencies in the concentration profiles only involve species *D* and *E*.

As the dimension of the kernel is not zero and the rank of the augmented matrix (*nc* = 4) is lower than the number of species (*ns* = 5), the pseudo-inverse of **C** cannot be computed.

Two options (among others) to allow computing the pseudo-inverse of **C** could be an additional experiment (*ne* = 2) with different initial concentrations (Strategy 4 again) or to define one uncoloured species (Strategy 1).

If Strategy (4) is applied once again, the initial concentrations to be varied must be selected amongst the species that have non-zero entries in the kernel, i.e. species *D* and *E*. An alternative is to define one uncoloured species (Strategy 1) amongst species *D* and *E*. Note that species *A*, *B* and *C* cannot be set uncoloured as their related concentrations are now linearly independent from the others. If, for example, species *E* is defined uncoloured, the linear combinations of true pure component spectra are given by Equation 4.20 with the independent rows of the augmented matrix (rows two, three and six as discussed in Section 4.4.3.1, and row seven corresponding to the additional experiment varying ratio from α to β).

$$\mathbf{\Delta} = \left(\mathbf{\Omega}_{\text{ind}}\Big|_{\varnothing/B}^{E/\varnothing}\right)^{-1}\mathbf{\Omega}_{\text{ind}}\Big|_{\varnothing/B}^{\varnothing/\varnothing} = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & k' \end{bmatrix} \end{array} \begin{matrix} \\ 'A' \\ 'B' \\ 'C' \\ 'D' \end{matrix} \tag{4.50}$$

With $k' = k_3/k_2$.

Species *A*, *B* and *C* are spectroscopically resolved (rows one, two and three) and the fitted component spectrum of the coloured species '*D*' is subject to linear combinations with the spectrum of the uncoloured species *E*.

4.5 Conclusion

The determination of a reaction mechanism based on spectroscopic kinetic investigations is an evolutionary process requiring a constant adaptation of the postulated kinetic model in order to comply with the main experimental observations made by chemists. When the underlying kinetic model is known (as it was assumed here in the presented examples), more attention has to be given to the determination of accurate related kinetic parameters, as this information is then used by chemical engineers in order to optimise the reaction conditions. For this two-fold task, kinetic hard-modelling is a particularly suitable technique. It allows the validation of the postulated model and of its related optimised kinetic parameters, on one hand by statistical means (e.g. goodness of the fit, analysis of the residuals or uncertainty estimation on the optimised parameters), but more importantly by comparing the optimised rate constants and the fitted component spectra with the experimenter's chemical knowledge or expectations.

For this purpose, we developed a novel method applicable to any type of bilinear spectroscopic data to be analysed by the kinetic hard-modelling technique proposed by Maeder and Zuberbühler. This method identifies in a systematic way, under ideal conditions, the minimum requirements in terms of mathematical pre-treatment (Strategy 1), a priori information (Strategy 2), or experimental design (strategies 3 and 4) in order to allow optimising the rate constants of a kinetic model. The method also predicts the conditions for which full spectral resolution can be expected. In case of partial spectral resolution (Strategy 1), the method determines the complex linear combinations of the true pure component spectra leading to the fitted (unresolved) component spectra.

This systematic method is based on the kernel of a time invariant augmented matrix $\boldsymbol{\Omega}$ that is pseudo-equivalent to the time variant concentration matrix **C** and covers rank deficiencies due to the stoichiometry (matrix **N**) and to the rate laws (matrix **E**). Compared to matrix **C**, setting up matrix $\boldsymbol{\Omega}$ is a simple task that can be quickly performed at the Matlab prompt, as it does not require any numerical integration of the differential equations. The kernel of matrix $\boldsymbol{\Omega}$ can be easily computed with Matlab's null command, based on Singular Value Decomposition (SVD).

Symbolic calculation of the kernel allows detecting particular stoichiometric conditions leading to spectral resolution of some species. For the common mechanism of Example 2 ($A + B \rightarrow C$, $A + C \rightarrow D$), we have shown, for example, that the fitted pure component spectra of species *C* and *D* are only fully resolved under stoichiometric ($c_{0,B}/c_{0,A} = 1$) and half-stoichiometric ($c_{0,B}/c_{0,A} = 1/2$) conditions respectively.

The method relies on the strict assumption that the rank deficiency in the measured spectroscopic data is due to a rank deficiency in the concentration matrix and not in the true pure component spectra of the absorbing species. The validity of this hypothesis depends on the spectroscopic nature of the signal. This assumption generally holds for mid-IR spectroscopy with a fairly high resolution and a low probability of linear dependent component spectra. For UV-vis spectroscopy, however, the absorption peaks are generally much broader and linear dependencies cannot be totally excluded. The method also assumes that all absorbing species contribute to the measured spectroscopic data, i.e. in time as well as in wavelength direction. In time direction, this assumption only depends on the technical capabilities of the instrument in terms of acquisition time, certain fast reactions requiring for example stopped-flow equipment or model reduction by applying a steady state approximation. In wavelength direction, the fulfilment of this assumption depends on the absorbing nature (absorptivities) of the reactive species with respect to the signal to noise ratio.

Although some of the underlying assumptions mentioned in Section 4.3.2 are rather idealised, knowing the optimal design to perform a kinetic analysis under ideal conditions is the first step to successfully complete this analysis under real conditions.

Acknowledgements

We thank Professor Daniel Kressner from the Seminar for Applied Mathematics at ETH Zurich for his mathematical advice.

4.6 Appendix

4.6.1 Reduction of time variant Equation 4.15 to time invariant Equation 4.16

In this section, we describe the different steps required to simplify time variant Equation 4.15, i.e. $\mathbf{\Omega}$ (3·*nt* x *ns*), into time invariant Equation 4.16, i.e. $\mathbf{\Omega}$ (*ns*+*nf*+1 x *ns*), while keeping the property of pseudo-equivalence with matrix **C**, i.e. $\mathbf{C} \sim \mathbf{\Omega}$. Note that if **C** and $\mathbf{\Omega}$ share the same kernel, they also share the same image, and their respective dimension, i.e. their defect and their rank, are the same between **C** and $\mathbf{\Omega}$.

The central point of this simplification is to apply row elementary operations on the time variant matrix $\mathbf{\Omega}$, e.g. left-multiplication by a diagonal or triangular matrix $\mathbf{\Theta}$, as these operations do not modify the matrix properties of $\mathbf{\Omega}$, i.e. its kernel and its image.

4.6.1.1 Elimination of diagonal matrices D and (I – D)

As **D** and (**I** – **D**) are invertible diagonal matrices, an appropriate diagonal matrix $\mathbf{\Theta}$ of dimensions (3·*nt* x *ns*) can be designed to remove **D** and (**I** – **D**) from $\mathbf{\Omega}$ without modifying its matrix properties.

$$\mathbf{C}(nt \text{ x } ns) \sim \mathbf{\Theta\Omega} = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{D}^{-1} & 0 \\ 0 & 0 & (\mathbf{I}-\mathbf{D})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{XN} \\ \mathbf{DC_0} \\ (\mathbf{I}-\mathbf{D})\mathbf{C_{dos}} \end{bmatrix} = \begin{bmatrix} \mathbf{XN} \\ \mathbf{C_0} \\ \mathbf{C_{dos}} \end{bmatrix} = \mathbf{\Omega}(3\,nt \text{ x } ns) \qquad (4.51)$$

4.6.1.2 Substitution of time variant C_0 and C_{dos} by time invariant c_0 and C_{in}

As $\mathbf{C_0}$ (*nt* x *ns*) and $\mathbf{C_{dos}}$ (*nt* x *ns*) are identical vertical stackings of $\mathbf{c_0}$ (1 x *ns*) and $\mathbf{C_{in}}$ (*nf* x *ns*) respectively, it is possible to multiply $\mathbf{C_0}$ and $\mathbf{C_{dos}}$ by an appropriate lower triangular matrix $\mathbf{\Theta}$ without affecting their respective matrix properties. For this, also note that removing rows only comprised by zeros does not change the properties of a matrix either.

$$\mathbf{C_0}(nt \text{ x } ns) \sim \mathbf{\Theta C_0} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{c_0} \\ \mathbf{c_0} \\ \mathbf{c_0} \\ \vdots \\ \mathbf{c_0} \end{bmatrix} = \begin{bmatrix} \mathbf{c_0} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = [\mathbf{c_0}] = \mathbf{c_0}(1 \text{ x } ns) \qquad (4.52)$$

$$\mathbf{C}_{\mathbf{dos}}\left(nt \text{ x } ns\right) \sim \mathbf{\Theta C}_{\mathbf{dos}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{c}_{\mathbf{in}\ 1,:} \\ \mathbf{c}_{\mathbf{in},\ 1,:} \\ \vdots \\ \mathbf{c}_{\mathbf{in}\ 2,:} \\ \mathbf{c}_{\mathbf{in}\ 2,:} \\ \mathbf{c}_{\mathbf{in},\ 2,:} \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \mathbf{c}_{\mathbf{in}\ 1,:} \\ 0 \\ \vdots \\ \mathbf{c}_{\mathbf{in}\ 2,:} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{\mathbf{in}\ 1,:} \\ \mathbf{c}_{\mathbf{in},\ 2,:} \\ \vdots \end{bmatrix} = \mathbf{C}_{\mathbf{in}}\left(nf \text{ x } ns\right) \tag{4.53}$$

The replacement of the time variant matrices $\mathbf{C_0}$ (*nt* x *ns*) and $\mathbf{C_{dos}}$ (*nt* x *ns*) by their respective time invariant vector $\mathbf{c_0}$ (1 x *ns*) and matrix $\mathbf{C_{in}}$ (*nf* x *ns*) leads to a matrix $\mathbf{\Omega}$ of dimensions (*nt*+*nf*+1 x *ns*).

$$\mathbf{C}\left(nt \text{ x } ns\right) \sim \begin{bmatrix} \mathbf{XN} \\ \mathbf{c_0} \\ \mathbf{C_{in}} \end{bmatrix} = \mathbf{\Omega}\left(nt + nf + 1 \text{ x } ns\right) \tag{4.54}$$

4.6.1.3 Pseudo-equivalence between the reaction extent and its derivative

The matrix of extents of reaction **X** (*nt* x *nr*) is calculated by numerical integration of a set of first order ordinary differential equations, i.e. involving the first derivative of **X**, denoted $\dot{\mathbf{X}}$ (*nt* x *nr*). Derivation of **X** (*nt* x *nr*) with respect to time **t** (*nt* x 1) leading to $\dot{\mathbf{X}}$ (*nt* x *nr*) can be described by a set of matrix multiplications involving a lower triangular matrix $\mathbf{\Theta}$ (*nt* x *nt*).

$$\dot{\mathbf{X}} = \mathrm{DIAG}\left(\mathbf{\Theta t}\right)^{-1} \mathbf{\Theta X} = \mathrm{DIAG}\left(\mathbf{\Theta} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ \vdots \\ t_{nt} \end{bmatrix} \right)^{-1} \mathbf{\Theta} \begin{bmatrix} \mathbf{x}_{1,:} \\ \mathbf{x}_{2,:} \\ \mathbf{x}_{3,:} \\ \mathbf{x}_{4,:} \\ \vdots \\ \mathbf{x}_{nt,:} \end{bmatrix} \quad \text{with} \quad \mathbf{\Theta} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{4.55}$$

As $\mathrm{DIAG}\left(\mathbf{\Theta t}\right)^{-1} \mathbf{\Theta}$ is a lower triangular matrix, it can be omitted without affecting the matrix properties of **X**. This implies that $\ker \dot{\mathbf{X}} = \left(\ker \mathbf{X}\right)\mathbf{T_{lin}}$, where $\mathbf{T_{lin}}$ is a linear transformation accounting for the rotational ambiguity of the kernel, and thus,

$$\dot{\mathbf{X}}\left(nt \text{ x } nr\right) \sim \mathbf{X}\left(nt \text{ x } nr\right) \tag{4.56}$$

4.6.1.4 Pseudo-equivalence between the reaction extent and the rate law

The system of first order ordinary differential equations leading to **X** can be elegantly written in matrix notation, introducing the matrix of rate laws denoted $\mathbf{\Phi}$ (*nt* x *nr*).

$$\dot{\mathbf{X}} = \mathbf{\Phi} - \mathrm{DIAG}(\mathbf{f})\mathrm{DIAG}(\mathbf{v})^{-1}\mathbf{X} \tag{4.57}$$

Using the definition of the kernel of $\dot{\mathbf{X}}$, i.e. $\dot{\mathbf{X}}(\ker \dot{\mathbf{X}}) = \mathbf{0}$, one can write:

$$\dot{\mathbf{X}}(\ker \dot{\mathbf{X}}) = \left(\mathbf{\Phi} - \mathrm{DIAG}(\mathbf{f})\mathrm{DIAG}(\mathbf{v})^{-1}\mathbf{X}\right)\ker \dot{\mathbf{X}} = \mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \mathrm{DIAG}(\mathbf{f})\mathrm{DIAG}(\mathbf{v})^{-1}\mathbf{X}(\ker \dot{\mathbf{X}}) = \mathbf{0} \tag{4.58}$$

As the product of DIAG(**f**) and DIAG(**v**)$^{-1}$ is a diagonal matrix, this left-multiplier can be omitted without affecting the matrix properties of **X**. Using the result of Section 4.6.1.3, i.e. $\ker \dot{\mathbf{X}} = (\ker \mathbf{X})\mathbf{T}_{\mathbf{lin}}$, and the definition of the kernel of **X**, i.e. $\mathbf{X}(\ker \mathbf{X}) = \mathbf{0}$, one can write:

$$\mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \mathbf{X}(\ker \dot{\mathbf{X}}) = \mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \mathbf{X}(\ker \mathbf{X})\mathbf{T}_{\mathbf{lin}} = \mathbf{\Phi}(\ker \dot{\mathbf{X}}) = \mathbf{0} \tag{4.59}$$

The last equality implies that $\ker \mathbf{\Phi} = (\ker \mathbf{X})\mathbf{T}_{\mathbf{lin}} = (\ker \dot{\mathbf{X}})\mathbf{T}'_{\mathbf{lin}}$, and thus:

$$\mathbf{C}(nt \text{ x } ns) \sim \begin{bmatrix} \mathbf{\Phi N} \\ \hline \mathbf{c_0} \\ \hline \mathbf{C_{in}} \end{bmatrix} = \mathbf{\Omega}(nt + nf + 1 \text{ x } ns) \tag{4.60}$$

4.6.1.5 Simplification of the time variant rate law into a time invariant rate law

The matrix of rate laws, $\mathbf{\Phi}$ (*nt* x *nr*), is calculated according to the first term on the right-hand side of Equation 4.6.

$$\mathbf{\Phi} = \begin{bmatrix} c_{1,1}^{e_{1,1}} c_{1,2}^{e_{1,2}} \cdots c_{1,\mathrm{ns}}^{e_{1,ns}} & c_{1,1}^{e_{2,1}} c_{1,2}^{e_{2,2}} \cdots c_{1,\mathrm{ns}}^{e_{2,ns}} & \cdots & c_{1,1}^{e_{\mathrm{nr},1}} c_{1,2}^{e_{\mathrm{nr},2}} \cdots c_{1,\mathrm{ns}}^{e_{\mathrm{nr},ns}} \\ c_{2,1}^{e_{1,1}} c_{2,2}^{e_{1,2}} \cdots c_{2,\mathrm{ns}}^{e_{1,ns}} & c_{2,1}^{e_{2,1}} c_{2,2}^{e_{2,2}} \cdots c_{2,\mathrm{ns}}^{e_{2,ns}} & \cdots & c_{2,1}^{e_{\mathrm{nr},1}} c_{2,2}^{e_{\mathrm{nr},2}} \cdots c_{2,\mathrm{ns}}^{e_{\mathrm{nr},ns}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\mathrm{nt},1}^{e_{1,1}} c_{\mathrm{nt},2}^{e_{1,2}} \cdots c_{\mathrm{nt},\mathrm{ns}}^{e_{1,ns}} & c_{\mathrm{nt},1}^{e_{2,1}} c_{\mathrm{nt},2}^{e_{2,2}} \cdots c_{\mathrm{nt},\mathrm{ns}}^{e_{2,ns}} & \cdots & c_{\mathrm{nt},1}^{e_{\mathrm{nr},1}} c_{\mathrm{nt},2}^{e_{\mathrm{nr},2}} \cdots c_{\mathrm{nt},\mathrm{ns}}^{e_{\mathrm{nr},ns}} \end{bmatrix} \mathrm{DIAG}(\mathbf{k}) \tag{4.61}$$

Matrix $\mathbf{\Phi}$ is based on a complex non-linear operation (see Equation 4.6) that formally requires defining three dimensional arrays of dimensions (*nt* x *ns* x *nr*), obtained by *nr*-times identical stacking matrix **C** (*nt* x *ns*) and *nt*-times identical stacking matrix **E** (*nr* x *ns*). An anti-clockwise rotation along the time axis (equivalent to a transposition of matrix **E**) and an unfolding in the time direction allows then reducing these three dimensional arrays into two dimensional arrays and the calculation of the matrix of rate laws, $\mathbf{\Phi}$, as presented just above.

Under the condition that rank deficiency in $\mathbf{\Phi}$ only occurs for strictly parallel elementary reactions whose rate laws have the same partial integer orders but different rate constants, the linear dependencies in the time variant $\mathbf{\Phi}$ are captured by the following time invariant matrix:

$$\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^{\mathrm{T}}}\,\mathrm{DIAG}\left(\mathbf{k}\right)\;\left(ns\;\mathrm{x}\;nr\right)\sim\mathbf{\Phi}\left(nt\;\mathrm{x}\;nr\right) \tag{4.62}$$

With μ being an arbitrary positive scalar different from 0 or 1, matrix $\mathbf{1}$ (*ns* x *nr*) being comprised of ones and the superscript $\bullet\mathbf{E}^{\mathrm{T}}$ representing the element-wise raise to the power of $\mathbf{E}^{\mathrm{T}}$, as also used in the calculation of the matrix of rate laws $\mathbf{\Phi}$.

Note, while the multiplication with DIAG(**k**) is not required to ensure the pseudo-equivalence in the above mentioned equation, it is convenient to see the analogy with the equation defining $\mathbf{\Phi}$, and also necessary to ensure strict pseudo-equivalence between $\mathbf{\Omega}$ and **C**.

A time invariant power base ($\mu\mathbf{1}$) is introduced to mimic the time variant power base comprised by the product of concentrations without affecting the matrix properties, in particular the kernel. Importantly, μ must not be equal to 0 or 1, as raising a base of 0 or 1 to the power of $\mathbf{E}^{\mathrm{T}}$, i.e. to the partial reaction orders, leads to the loss of the information contained in $\mathbf{E}^{\mathrm{T}}$ and thus breaks the pseudo-equivalence between $\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^{\mathrm{T}}}$ and $\mathbf{\Phi}$.

The pseudo-equivalence between $\mathbf{\Phi}$ and $\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^{\mathrm{T}}}$ finally allows to write a time invariant $\mathbf{\Omega}$ (*ns*+*nf*+1 x *ns*), as presented in Equation 4.16, that is pseudo-equivalent to the time variant $\mathbf{\Omega}$ (3·*nt* x *ns*) of Equation 4.15 and to the matrix **C** (*nt* x *ns*).

$$\mathbf{\Omega}\left(ns+nf+1\;\mathrm{x}\;ns\right)=\begin{bmatrix}\left(\mu\mathbf{1}\right)^{\bullet\mathbf{E}^{\mathrm{T}}}\,\mathrm{DIAG}\left(\mathbf{k}\right)\mathbf{N}\\ \hline \dfrac{\mathbf{c_0}}{\mathbf{C_{in}}}\end{bmatrix}\sim\mathbf{C}\left(nt\;\mathrm{x}\;ns\right) \tag{4.63}$$

4.6.2 Matlab example for the setup of $\boldsymbol{\Omega}$ applied to the kinetic scheme of Figure 4.1 and the calculation of its kernel under batch conditions

Table 4.4: Setup and calculation of the kernel of $\boldsymbol{\Omega}$ under batch conditions for the kinetic scheme of Figure 4.1.

Particular solution with $\alpha = c_{0,B}/c_{0,A} = 1$	General solution
>> mu = 0.33;	>> syms c0A alpha mu k1 k2
>> N = [-1 -1 1 0; -1 0 -1 1];	>> N = [-1 -1 1 0; -1 0 -1 1];
>> E = [1 1 0 0; 1 0 1 0];	>> E = [1 1 0 0; 1 0 1 0];
>> k = [1, 2];	>> k = [k1, k2];
>> c0 = [1, 1, 0, 0];	>> c0 = [c0A, alpha*c0A, 0, 0];
>> one = ones(size(E'));	>> one = ones(size(E'));
>> omega = [(mu*one).^(E')*diag(k)*N; c0];	>> omega = [(mu*one).^(E')*diag(k)*N; c0];
>> null(omega)	>> null(omega)
ans =	ans =
0.5774	-alpha
-0.5774	1
0.0000	1-alpha
0.5774	1-2*alpha

4.6.3 Comparison of ker $\boldsymbol{\Omega}$ with ker C for the kinetic scheme of Equation 4.39 under batch conditions

Assuming experimental conditions $c_{0,A}$ = 1 mol·L^{-1}, α = 1 and rate constants **k** = [0.5, 0.4, 0.3], both ker **C** (5 x 2) and ker $\boldsymbol{\Omega}$ (5 x 2) can be calculated using Singular Value Decomposition (SVD).

$$\ker \mathbf{C} = \begin{bmatrix} -0.5015 & 0.0322 \\ 0.5015 & -0.0322 \\ 0 & 0 \\ -0.5232 & 0.6349 \\ -0.4726 & -0.7713 \end{bmatrix}, \ \ker \boldsymbol{\Omega} = \begin{bmatrix} -0.4986 & 0.0625 \\ 0.4986 & -0.0625 \\ 0 & 0 \\ -0.6331 & -0.5254 \\ -0.3194 & 0.8463 \end{bmatrix} \tag{4.64}$$

Where values are rounded to the fourth digit.

The kernel is a basis for a vector space and thus any linear combination of this basis, i.e. linear transformation, spans the same vector space (rotational ambiguity).

In particular, Singular Value Decomposition (SVD) delivers a set of orthogonal vectors and in this case this linear transformation is called a rotation. The agreement between ker **C** and ker **Ω** is demonstrated by showing that ker **Ω** can be obtained by linear combinations of ker **C**.

$$\ker\ \mathbf{\Omega} = (\ker\ \mathbf{C})\mathbf{T}_{\mathbf{lin}} \quad \text{with } \mathbf{T}_{\mathbf{lin}} = (\ker \mathbf{C})^{+} \ker\ \mathbf{\Omega} = \begin{bmatrix} 0.9822 & -0.1878 \\ -0.1878 & -0.9822 \end{bmatrix} \tag{4.65}$$

The analytical kernel of $\mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$, as calculated in Table 4.3, is also in agreement with the kernel of **C**, as a rotation of ker **C** leads to $\ker\ \mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing}$.

$$\ker\ \mathbf{\Omega}\Big|_{\varnothing/\varnothing}^{\varnothing/\varnothing} = \begin{bmatrix} -\alpha & 0 \\ 1 & 0 \\ 1-\alpha & 0 \\ (1+k') & (1-2\alpha) \quad -k' \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 0 \\ -1.75 & -0.75 \\ 0 & 1 \end{bmatrix} = (\ker \mathbf{C})\mathbf{T}_{\mathbf{lin}} \quad \text{with } \mathbf{T}_{\mathbf{lin}} = \begin{bmatrix} 1.9187 & -0.0802 \\ -1.1756 & -1.2474 \end{bmatrix} \tag{4.66}$$

With $\alpha = 1$ and $k' = k_3/k_2 = 0.75$.

4.6.4 Kernel of Ω when strategies (1) to (4) are applied to the kinetic scheme of Equation 4.39

Table 4.5: Kernel of **Ω** (for the kinetic scheme of Equation 4.39) when one column is removed from **Ω** (Strategy 1 or 2) or when one row is added to **Ω** (Strategy 3 or 4).

Species	Species *A* used in Strategy (1, 2, 3 or 4)	Species *B* used in Strategy (1, 2, 3 or 4)	Species *C* used in Strategy (1, 2, 3 or 4)		Species *D* used in Strategy (1, 2, 3 or 4)	Species *D* used in Strategy (1, 2, 3 or 4)
A	n/a or 0 [1)]	0	0	$-\varepsilon(\alpha)$ [3)]	$-\alpha$	$-\alpha$
B	0	n/a or 0 [1)]	0	$\varepsilon(\alpha)$	1	1
C	0	0	n/a or 0 [1)]	n/a or 0 [1)]	$(1-\alpha)$	$(1-\alpha)$
D	$-k'$ [2)]	$-k'$	$-k'$	$-\varepsilon(\alpha)(1+k')$	n/a or 0 [1)]	$(1+k')(1-2\alpha)$
E	1	1	1	0	$(1+(k')^{-1})(1-2\alpha)$	n/a or 0 [1)]

[1)] n/a if Strategy (1) or (2) is applied, 0 if Strategy (3) or (4) is used.

[2)] $k' = k_3/k_2$.

[3)] $\varepsilon = 0$ for $\alpha \neq 1$, $\varepsilon = 1$ for $\alpha = 1$.

4.6.5 Prediction of linear dependencies in the reaction extents and implications on the kinetic hard-modelling of calorimetric data

According to Section 2.1.1 and to Equations 2.4 and 2.5, the reaction heat power $\dot{\mathbf{q}}_r$ (*nt* x 1) can be expressed as the product of the reaction volume **v** (*nt* x 1), the temporal derivative of the extent of reaction expressed in moles $\dot{\boldsymbol{\xi}}$ (*nt* x *nr*) and the reaction enthalpies $\boldsymbol{\Delta}_r\mathbf{H}$ (*nr* x 1).

$$\dot{\mathbf{q}}_r = \dot{\boldsymbol{\xi}}\left(-\boldsymbol{\Delta}_r\mathbf{H}\right) \qquad \text{with} \quad \dot{\boldsymbol{\xi}} = \mathrm{DIAG}(\mathbf{v})\,\boldsymbol{\Phi} \tag{4.67}$$

Reaction enthalpies $\boldsymbol{\Delta}_r\mathbf{H}$ can be eliminated from the non-linear optimisation and replaced by their linear least-squares estimates (see analogy with Equation 4.9 for spectroscopic data).

$$\boldsymbol{\Delta}_r\mathbf{H} = -\dot{\boldsymbol{\xi}}^{+}\dot{\mathbf{q}}_r \tag{4.68}$$

This linear regression is only possible if $\dot{\boldsymbol{\xi}}$ is of full rank. The derivatives of the reaction extents can, however, be rank deficient for strictly parallel reactions, i.e. for two or more rate laws with the same partial orders (identical rows in matrix **E**) just differing by their rate constants. Two strategies can be applied to circumvent rank deficiencies in matrix $\dot{\boldsymbol{\xi}}$: defining non-contributing reaction extents (Strategy 1) or including some independently known reaction enthalpies to the analysis (Strategy 2). Note that defining non-contributing reactions, although they are in fact contributing to $\dot{\mathbf{q}}_r$, has the drawback that the missing calorimetric contributions are transferred into the enthalpies of the contributing reactions. This is similar as the effect described for spectroscopy when uncoloured species are defined. Also, note that dosing or performing a second order global analysis (Strategy 3 and Strategy 4) cannot be used to break linear dependencies in the reaction extents of irreversible reactions, as these varied experimental conditions have the same impact on each rate law of strictly parallel reactions.

Strategies (1) and (2) can be described in a similar way as for spectroscopic data, except that the predictions have to be based on a time invariant matrix $\boldsymbol{\Omega}$ that is pseudo-equivalent to **X**.

$$\mathbf{\Omega}\ (ns \text{ x } nr) = (\mu \mathbf{1})^{\bullet \mathbf{E}^T}\ \text{DIAG}(\mathbf{k}) \sim \mathbf{X}\ (nt \text{ x } nr) \tag{4.69}$$

Linear dependencies in **X** can be identified using ker $\mathbf{\Omega}$ or more easily by identifying identical rows in matrix **E**. Effects of Strategies (1) and (2) are shown with two examples including two and three strictly parallel reactions.

4.6.5.1 Application of Strategy (1) – defining non-contributing reactions

Strategy (1) is applied to a kinetic scheme comprised of two strictly parallel reactions.

$$\begin{aligned} A + B &\xrightarrow{k_1} C \\ A + B &\xrightarrow{k_2} D \end{aligned} \qquad \text{with } \mathbf{E} = \begin{matrix} A & B & C & D \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \tag{4.70}$$

The heat signal for this kinetic scheme can be calculated according to the individual reaction extent derivatives and enthalpies.

$$\dot{\mathbf{q}}_\mathbf{r} = \dot{\boldsymbol{\xi}}(-\mathbf{\Delta}_\mathbf{r}\mathbf{H}) = \dot{\boldsymbol{\xi}}_{t,1}(-\Delta_r H_1) + \dot{\boldsymbol{\xi}}_{t,2}(-\Delta_r H_2) \tag{4.71}$$

With $\dot{\boldsymbol{\xi}}_{t,1} = \mathbf{v}_t k_1 \mathbf{c}_{t,A} \mathbf{c}_{t,B}$ and $\dot{\boldsymbol{\xi}}_{t,2} = \mathbf{v}_t k_2 \mathbf{c}_{t,A} \mathbf{c}_{t,B}$. Note that, $\dot{\boldsymbol{\xi}} = [\dot{\boldsymbol{\xi}}_{t,1}, \dot{\boldsymbol{\xi}}_{t,2}]$ is rank deficient for this mechanism as $\dot{\boldsymbol{\xi}}_{t,2} = \left(k_2 k_1^{-1}\right)\dot{\boldsymbol{\xi}}_{t,1}$. Thus, $\dot{\boldsymbol{\xi}}^+$ cannot be computed.

When Strategy (1) is used to circumvent the rank-deficiency of $\dot{\boldsymbol{\xi}}$, one of the linearly dependent reactions has to be defined as non-contributing. If, for example, the second reaction is defined as non-contributing, the fitted enthalpy $\Delta_r H_c$ of the contributing reaction is a linear combination of the true enthalpies of the first and second reaction. Linear coefficients can be determined based on the principle of energy conservation.

$$\dot{\boldsymbol{\xi}}_{t,1}(-\Delta_r H_c) = \dot{\boldsymbol{\xi}}_{t,1}(-\Delta_r H_1) + \dot{\boldsymbol{\xi}}_{t,2}(-\Delta_r H_2) \tag{4.72}$$

Using the relationship between $\dot{\boldsymbol{\xi}}_{t,1}$ and $\dot{\boldsymbol{\xi}}_{t,2}$, i.e. $\dot{\boldsymbol{\xi}}_{t,2} = \left(k_2 k_1^{-1}\right)\dot{\boldsymbol{\xi}}_{t,1}$, the energy balance can be rearranged in order to get an expression for the fitted enthalpy $\Delta_r H_c$.

$$\Delta_r H_c = \Delta_r H_1 + k_2 k_1^{-1} \Delta_r H_2 = \mathbf{\Delta}\left(\mathbf{\Delta_r H}\right) \qquad \text{with } \mathbf{\Delta} = \left[1 \;\; k_2 k_1^{-1}\right] \tag{4.73}$$

This result indicates that only one enthalpy can be fitted for a set of strictly parallel reactions. The fitted reaction enthalpy is a linear combination of the individual reaction enthalpies with linear coefficients being ratios between the individual rate constants of the strictly parallel reactions (k_1, k_2, k_3 …) and the rate constant of the contributing reaction (here k_1), i.e. $\mathbf{\Delta}\ \left(nr - nu_r \text{ x } nr\right) = \left[1 \;\; k_2 k_1^{-1} \;\; k_3 k_1^{-1} \;\; \ldots\right]$. In this notation, nu_r denotes the number of non-contributing reactions. Note that the linear coefficients given in $\mathbf{\Delta}$ do not depend on the kinetic order of the strictly parallel reactions.

4.6.5.2 Application of Strategy (2) – including known enthalpies into the analysis

When Strategy (2) is used to circumvent linear dependencies in $\dot{\boldsymbol{\xi}}$, *nke* = rank ($\dot{\boldsymbol{\xi}}$) reaction enthalpies have to be included in the analysis in order to reach full calorimetric resolution, *nke* denoting the number of known enthalpies that are provided. If *nke* is lower than the rank of $\dot{\boldsymbol{\xi}}$, Strategies (1) and (2) have to be combined and a linear combination of the remaining reaction enthalpies is fitted (partial resolution of the enthalpies). In such case, linear coefficients in matrix $\mathbf{\Delta}$ (*nr–nu*$_r$*–nke* x *nr–nke*) are given by the ratios between the individual rate constants of the strictly parallel reactions, for which no enthalpies have been provided, and the rate constant of the contributing reaction.

As an example, consider a kinetic scheme comprised of three strictly parallel reactions.

$$\begin{aligned} A + B &\xrightarrow{k_1} C \\ A + B &\xrightarrow{k_2} D \\ A + B &\xrightarrow{k_3} E \end{aligned} \qquad \text{with } \mathbf{E} = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{4.74}$$

If the enthalpy of the third reaction, i.e. $\Delta_r H_3$, is independently known, it can be included into the analysis. However, as the rank of $\dot{\boldsymbol{\xi}}$ is only one (see matrix **E**), Strategy (1) has also to be applied and one of the two remaining reactions has to be defined as non-contributing.

If the first reaction is defined as non-contributing, the fitted enthalpy of the second reaction is a linear combination of the enthalpies of the first and second reactions. For this example, matrix $\mathbf{\Delta}$ has dimensions (*nr*–*nu*$_r$–*nke* x *nr*–*nke*) = (3–1–1 x 3–1) = (1 x 2).

$$\Delta_r H_c = k_1 k_2^{-1} \Delta_r H_1 + \Delta_r H_2 = \mathbf{\Delta}\left(\mathbf{\Delta}_\mathbf{r}\mathbf{H}\right) \qquad \text{with } \mathbf{\Delta} = \left[k_1 k_2^{-1} \;\; 1\right] \tag{4.75}$$

4.7 References

[1] P. Geladi, B.R. Kowalski, Anal. Chim. Acta, 185 (1986), 1-17.
[2] M. Maeder, A.D. Zuberbühler, Anal. Chem., 62 (1990), 2220-2224.
[3] E. Sanchez, B.R. Kowalski, J. Chemom., 4 (1990), 29-45.
[4] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, Comput. Chem. Eng., 20 (1996), 1451-1465.
[5] R. Bro, Chemom. Intell. Lab. Syst., 38 (1997), 149-171.
[6] E.R. Malinowski, Factor Analysis in Chemistry, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
[7] M. Maeder, Y.M. Neuhold, Chapter 7 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 218-256.
[8] R. Tauler, A. de Juan, Chapter 11 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 417-474.
[9] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, Trac-Trends Anal. Chem., 26 (2007), 752-765.
[10] M. Maeder, Y.M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, NL, 2007.
[11] E. Furusjö, L.G. Danielsson, Anal. Chim. Acta, 373 (1998), 83-94.
[12] S. Bijlsma, D.J. Louwerse, A.K. Smilde, AIChE J., 44 (1998), 2713-2723.
[13] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst., 59 (2001), 19-31.
[14] G. Puxty, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 81 (2006), 149-164.
[15] M. Amrhein, Reaction and flow variants/invariants for the analysis of chemical reaction data, Doctoral Thesis 1861, EPF Lausanne, Switzerland, 1998.
[16] V.M. Taavitsainen, H. Haario, J. Chemom., 15 (2001), 215-239.
[17] V.M. Taavitsainen, H. Haario, M. Laine, J. Chemom., 17 (2003), 140-150.
[18] Y.M. Neuhold, M. Maeder, J. Chemom., 16 (2002), 218-227.
[19] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta, 396 (1999), 231-240.
[20] R.N. Cochran, F.H. Horne, Anal. Chem., 49 (1977), 846-853.
[21] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemom. Intell. Lab. Syst., 33 (1996), 17-33.
[22] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurenczy, M. Maeder, A.E. Merbach, A.D. Zuberbühler, Anal. Chim. Acta, 298 (1994), 193-201.
[23] R. Dyson, M. Maeder, Y.M. Neuhold, G. Puxty, Anal. Chim. Acta, 490 (2003), 99-108.
[24] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst., 81 (2006), 82-93.
[25] J.J. Kankare, Anal. Chem., 42 (1970), 1322-&.
[26] M. Garland, E. Visser, P. Terwiesch, D.W.T. Rippin, Anal. Chim. Acta, 351 (1997), 337-358.
[27] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemom. Intell. Lab. Syst., 46 (1999), 249-264.
[28] R. Dyson, M. Maeder, G. Puxty, Y.M. Neuhold, Inorg. React. Mech., 5 (2003), 39-46.
[29] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C++ - The art of Scientific Computing, Second Edition, Cambridge University Press, New York, USA, 2005.
[30] A.E. Croce, J. Chem. Educ., 79 (2002), 506-509.
[31] H. Maskill, Educ. Chem., 21 (1984), 122-123.

[32] J. Billeter, Y.M. Neuhold, L.L. Simon, G. Puxty, K. Hungerbühler, Chemom. Intell. Lab. Syst., 93 (2008), 120-131.
[33] D.C. Lay, Linear Algebra and Its Applications, Third Edition, Addison Wesley Higher Education, Boston, USA, 2003.
[34] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007; http://www.mathworks.com
[35] M. Maeder, Y.M. Neuhold, A. Olsen, G. Puxty, R. Dyson, A. Zilian, Anal. Chim. Acta, 464 (2002), 249-259.

CHAPTER 5

Kinetic hard-modelling and spectral validation of rank deficient spectroscopic data: a case study

This chapter was published as an article under the reference:

J. Billeter, Y.-M. Neuhold, K. Hungerbühler
Chemometrics and Intelligent Laboratory Systems, 98 (2009), 213-226
http://dx.doi.org/10.1016/j.chemolab.2009.07.002

* * *

5.1 Abstract

In this case study, we apply a recently developed method to systematically predict the linear dependencies in concentration profiles and identify minimum requirements to enable optimisation of rate constants and pure component spectra via direct multivariate kinetic hard-modelling of spectroscopic data. This systematic method was applied to the rank-deficient acid catalysed reaction of benzophenone with phenylhydrazine in THF. Various experimental conditions (different dosing and initial concentrations) and data treatments (defining uncoloured species, including known component spectra into the analysis) were considered. For all these conditions, the kinetic mechanism of this condensation reaction was successfully validated by the agreement between fitted and independently measured mid-IR and UV-vis pure component spectra and by the highly reproducible fitted rate constants. This case study particularly demonstrated the value of the direct spectral fitting as a tool for the validation of rank-deficient kinetic mechanisms, as inherent contributions within the fitted component spectra, due to the definition of uncoloured species, can be systematically addressed.

5.2 Introduction

Various chemometric methods capable to analyse time dependent multivariate data measured by spectroscopic techniques have been introduced in recent years [1-10]. Amongst these chemometric methods, kinetic hard-modelling, based on a hard model (the rate law), can be used to directly determine the kinetic parameters (e.g. rate constants) of chemical reactions [2,8,10-14].

In spectroscopy, multivariate data, i.e. time and wavelength/wavenumber resolved absorbance measurements, can be decomposed according to Beer's law into the concentration profiles and the molar spectra of the pure components. In kinetic hard-modelling these concentration profiles are calculated by numerical integration of the rate laws describing the postulated kinetic hard model.

Maeder and Zuberbühler suggested to eliminate the pure component spectra from the non-linear optimisation and to linearly estimate them at each iteration by regressing the spectroscopic data on the concentration matrix via its pseudo inverse [2]. This latter method, used in the present article and sometimes referred to as kinetic hard-modelling by implicit direct calibration [15,16], allows the validation of the rate law by comparing the estimated pure component spectra with independently measured ones, or, if not fully accessible, at least with expected peaks.

For some kinetic models and experimental conditions, however, the concentration matrix is rank deficient and the pure component spectra cannot be computed, as the linear regression step cannot be performed. Different solutions have been proposed in order to circumvent this rank deficiency problem and to allow the fitting of pure component spectra: defining some absorbing species as uncoloured (Strategy 1) [10,17], including some independently known component spectra to the analysis (Strategy 2) [10,18], dosing one or more species (Strategy 3) [17] or varying some initial concentrations and analysing simultaneously the resulting experiments (Strategy 4) [14,19,20]. This last strategy is referred to as second order global analysis or 3-way analysis.

In a recent article [17] (Chapter 4), we have introduced a method for the systematic experimental and data analytical design of bi-linear spectroscopic kinetic measurements that allows identifying the species to be incorporated in strategies (1) to (4) to obtain partial or full spectral resolution. When only partial resolution is possible (Strategy 1), we have also presented a method for the calculation of the linear combinations in the fitted component spectra, i.e. the coefficients by which the true (resolved) pure spectra need to be weighted in order to obtain the reduced set of fitted (unresolved) component spectra corresponding to the species defined as coloured. Theoretical background of this approach was introduced and applied to simulated data based on various kinetic models.

In the present case study, we apply this systematic method to sets of kinetic data, measured by mid-IR and UV-vis spectroscopy simultaneously, recorded during the course of the reaction of benzophenone with phenylhydrazine in THF catalysed by acetic acid (see Figure 5.1). The kinetic mechanism of this reaction has been discussed in previous articles [21,22] and is used here as a test reaction to experimentally validate the theoretical concepts presented previously [17]. Strategies (1) to (4) are applied to this reaction in order to break the rank deficiency in the concentration matrix without distorting the calculated rate constants. As a consequence of Strategy (1), linear dependencies in the concentration profiles translate into the fitted component spectra, and are compared to those theoretically predicted by our method. The underlying kinetic model is spectroscopically validated using different experimental conditions, e.g. dosing different species, via the reproducibility of the fitted rate constants and the accuracy of the fitted component spectra.

5.3 Theoretical background

In spectroscopy, Beer's law, i.e. **Y** = **CA** + **R**, is used to decompose a measured absorbance signal **Y** (*nt* x *nw*) into the product of the concentrations **C** (*nt* x *ns*) and the molar spectra **A** (*ns* x *nw*) of the pure components. Deviations from the product of **C** and **A** are captured in a matrix **R** (*nt* x *nw*) of residuals. Matrix dimensions are defined using *nt* as the number of reaction times, *nw* as the number of wavelengths and *ns* as the total number of species. For details regarding the notation, we refer to a previous publication introducing the mathematical concepts of our approach [17]. In the following, we provide a short summary.

Kinetic hard-modelling uses the kinetic rate law to define a system of ordinary differential equations (ODE) that depend on kinetic parameters, e.g. rate constants. This chemical model can be uniquely identified using the matrix of reactants coefficients, **E** (*nr* x *ns*), and the matrix of stoichiometry, **N** (*nr* x *ns*), for all species involved in the *nr* elementary reaction steps [23]. The matrix of concentration profiles **C** can then be calculated for a given set of rate constants **k** (1 x *nr*) by numerical integration of the ODE, using initial concentrations, $\mathbf{c_0}$ (1 x *ns*), matrices **E** and **N**, as well as dosing conditions, i.e. dosing rates **f** (*nt* x 1) and concentrations $\mathbf{C_{in}}$ (*nf* x *ns*) of the *nf* dosing steps [14,24]. Note that in this somewhat simplified approach, changes in density or activity are not taken into account.

In direct kinetic hard-modelling of spectroscopic data, originally proposed by Maeder and Zuberbühler [2], the product of the integrated concentration profiles **C** and of the pure component spectra **A** is compared to the measured data matrix **Y**, and results in the residuals $\mathbf{R} = \mathbf{Y} - \mathbf{CA}$, capturing the differences between the measured and the modelled absorbances. In the least-squares analysis, the sum of all squared residuals **R** (*nt* x *nw*) is used as the objective function to be minimised by iteratively optimising the rate constants **k**. For this article, the Newton-Gauss-Levenberg/Marquardt algorithm (NGL/M) was used to solve this non-linear regression [2,10,24-27]. This gradient-based method allows estimating the uncertainties in the optimised rate constants from the variance/covariance matrix, including the propagation of errors, such as the uncertainties in initial concentrations and in dosing rate, as discussed previously [21].

As **A** is comprised of linear parameters only, it can be eliminated from the non-linear optimisation and replaced by its linear least-squares estimate [10,14].

$$\mathbf{A} = \mathbf{C}^{+}\mathbf{Y} = \left(\mathbf{C}^{\mathsf{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathsf{T}}\mathbf{Y} \tag{5.1}$$

Equation 5.1 is only applicable if **C** is of full rank. If **C** is rank deficient, one or more strategies (1 to 4) need to be used in order to allow the computation of **A** by linear regression. The species to include in these four strategies can be identified from the linear dependencies in the *ns* concentration profiles of **C**, obtained by the kernel of **C**, denoted ker **C**.

The kernel of **C** is defined as the vector basis spanned by the vectors forming the null space **0** when left-multiplied by **C**, i.e. **C** (ker **C**) = **0** [28]. Note, if the kernel is not empty, it also defines the mass balance (closure) of the chemical system. The analogy between kernel and mass balance is demonstrated in Sections 5.7.6 and 5.7.7 (Appendices).

Due to its definition, the kernel inherently involves a rotational ambiguity and thus any linear combinations of the kernel also span a basis for this vector space. Importantly, a row only comprised by zeros in ker **C** indicates that the corresponding concentration profile is linearly independent from the others.

To predict the linear dependencies in the concentration profiles without numerical integration, we have introduced a time invariant augmented matrix **Ω** of general dimensions (*ns*+*nf*+*ne*+1 x *ns*) having the same kernel as the time variant matrix of concentrations **C** (*nt* x *ns*) [17]. In this notation, (*ne*+1) defines the number of simultaneously analysed experiments. In Equation 5.2, matrix **Ω** is presented for the case of one single batch experiment (*nf* = 0 and *ne* = 0), i.e. **Ω** has dimensions (*ns*+1 x *ns*).

$$\boldsymbol{\Omega} = \begin{bmatrix} \dfrac{(\mu \mathbf{1})^{\bullet \mathbf{E}^T} \, \mathrm{DIAG}(\mathbf{k}) \mathbf{N}}{\mathbf{c_0}} \end{bmatrix} \tag{5.2}$$

With μ being an arbitrary positive scalar different from 0 and 1, matrix **1** (*ns* x *nr*) comprised of ones only and the superscript $\bullet\mathbf{E}^T$ representing the element-wise raise to the power of $\mathbf{E}^T$.

The prediction for the linear dependencies in **C** and in the fitted component spectra **A**, based on the analysis of **Ω**, assumes that (i) the kinetic hard-model is correct, (ii) no reactions are identical, (iii) all species initiating the reactions are present initially or are dosed, (iv) the true pure component spectra are linearly independent, (v) each provided pure component spectrum is correct, (vi) all species absorb or the truly non-absorbing species are known and defined such that they do not contribute to Beer's law, and (vii) all reaction steps defined by the kinetic hard-model can be observed in **Y**.

5.3.1 Strategy (1) – defining uncoloured species

When Strategy (1) is used to circumvent linear dependencies in the concentration profiles, Beer's law is reduced by defining (*ns*–*nu*) coloured and *nu* uncoloured species, and by eliminating from the time invariant matrix $\mathbf{\Omega}$ (and $\mathbf{C}$) the columns corresponding to these *nu* uncoloured species. This leads to reduced matrices $\mathbf{\Omega}$ and $\mathbf{C_c}$ of dimensions (*ns*+1 x *ns*–*nu*) and (*nt* x *ns*–*nu*) respectively. As a consequence of this reduction, the fitted component spectra of the coloured species $\mathbf{A_c}$ (*ns*–*nu* x *nw*) are comprised of linear combinations of the true (resolved) pure spectra $\mathbf{A}$ (*ns* x *nw*). To obtain the coefficients of these linear combinations, the linear dependencies in the rows of the non-reduced matrix $\mathbf{\Omega}$ (*ns*+1 x *ns*) have to be eliminated, leading to a matrix $\mathbf{\Omega_{ind}}$ of dimensions (nr_i+1 x *ns*), where nr_i denotes the number of independent reactions, and $\ker \mathbf{\Omega}_{\mathbf{ind}}^{\mathrm{T}} = \mathbf{0}$. Based on the coloured species and the linearly independent reactions, the coefficients for the linear combinations of the true pure component spectra, $\mathbf{\Delta}$ (*ns*–*nu* x *ns*), can be calculated as follows:

$$\mathbf{A_c} = \mathbf{\Delta}\,\mathbf{A} \quad \text{with} \quad \mathbf{\Delta} = \left(\mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\ \text{coloured species}}} \right)^{-1} \mathbf{\Omega_{ind}}\Big|_{\substack{\text{comprised of}\\ \text{all species}}} \tag{5.3}$$

5.3.2 Strategy (2) – including known spectra into the analysis

In this case, *nks* pure component spectra are included into the kinetic analysis in order to avoid linear dependencies in the concentration profiles. Here, the corresponding *nks* columns of $\mathbf{\Omega}$ and $\mathbf{C}$ are removed, leading to a reduced time-invariant matrix $\mathbf{\Omega}$ of dimensions (*ns*+1 x *ns*–*nks*) and the corresponding reduced matrix of concentrations $\mathbf{C_{uk}}$ of dimensions (*nt* x *ns*–*nks*).

5.3.3 Strategy (3) – dosing one or more species

This strategy employs *nf* successive dosing steps, such that the rank of $\mathbf{\Omega}$ (and $\mathbf{C}$) becomes augmented. Thus matrix $\mathbf{\Omega}$ is expanded to dimensions (*ns*+*nf*+1 x *ns*) with the appropriate rows corresponding to the dosed concentrations of each dosing step. The dosed concentrations are collected in matrix $\mathbf{C_{in}}$ (*nf* x *ns*).

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\ \mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c}_0} \\ \hline \mathbf{C}_{in} \end{bmatrix} \tag{5.4}$$

5.3.4 Strategy (4) – performing a second order global analysis

Strategy (4), also known as second order global analysis, is used to augment the rank of $\mathbf{\Omega}$ (and $\mathbf{C}$) by concatenating multiple kinetic experiments performed under various initial concentrations. When common pure component spectra are fitted (global mode), matrix $\mathbf{\Omega}$ is expanded to dimensions (*ns*+*ne*+1 x *ns*) by adding *ne* rows corresponding to the different initial concentrations of the *ne* additional experiments, which are collected in matrix $\mathbf{C}_0^{ne}$ (*ne* x *ns*).

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet\mathbf{E}^T}\ \mathrm{DIAG}(\mathbf{k})\mathbf{N}}{\mathbf{c}_0} \\ \hline \mathbf{C}_0^{ne} \end{bmatrix} \tag{5.5}$$

5.4 Experiments

The reaction of benzophenone (species *B*) with phenylhydrazine (*P*) in THF catalysed by acetic acid (*Aa*) can be monitored in mid-IR and UV-vis spectral ranges. Products of the reaction are benzophenone-phenylhydrazone (*BP*), water and the regenerated catalyst (*Aa*). This model reaction has been extensively described for acetic acid being dosed into benzophenone and phenylhydrazine [21,22]. In a recent paper (Chapter 3), we have also shown that this condensation reaction and the optimisation of its rate constant are highly reproducible [21], thus being an ideal case study to validate our systematic method to treat kinetic rank deficiencies.

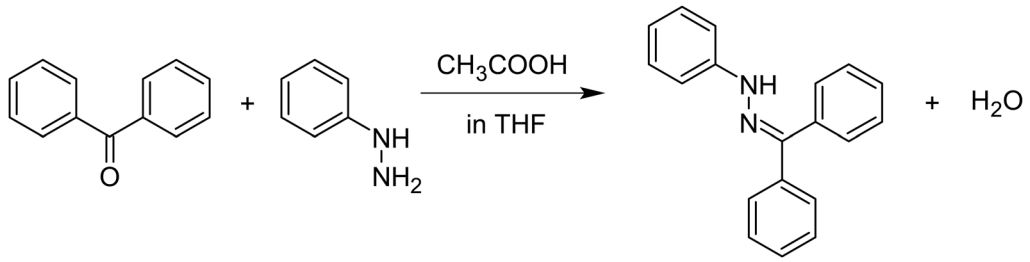

Figure 5.1: Reaction of benzophenone (*B*) with phenylhydrazine (*P*) catalysed by acetic acid (*Aa*) forming benzophenone-phenylhydrazone (*BP*) and water.

5.4.1 Instruments

Experiments were carried out in the Combined Reaction Calorimeter (CRC.v4), a small-scale reaction calorimeter working under isothermal conditions and combining the principle of power compensation and heat balance [29]. As the reaction does not produce significant amounts of heat, the calorimetric heat signal was disregarded and the well-defined environment of the CRC.v4 was only used to maintain a constant temperature of 25 ($\pm$ 0.04) °C for all experiments. For a detailed description of the reactor, we refer to [29].

Mid-IR signals were followed by ATR FT-IR spectroscopy (ReactIR 4000, K-4 conduit and ATR crystal from Mettler Toledo) in the range of 1200 to 1650 cm^{-1} at a resolution of 4 cm^{-1}. UV-vis signals were monitored by ATR UV-vis spectroscopy (Cary 50 from Varian coupled to an ATR dip-probe from Hellma) between 240 and 400 nm at 1 nm resolution. UV-vis and mid-IR signals were recorded simultaneously every minute during 150 minutes against THF as the reference background.

5.4.2 Sample preparation

A solution of phenylhydrazine in THF (Acros Organics for analysis, certified 99.99%) was prepared by weighing 40.003 g of phenylhydrazine (Aldrich-Fine Chemicals, certified 99.2%) into a volumetric flask and filling up to 250 mL with THF. A benzophenone solution was prepared by weighing 36.261 g of benzophenone (Fluka purum, certified 99.8%) and filling up to 100 mL with THF. This sample preparation led to stock concentrations of 1.468 ($\pm$ 0.001) $mol \cdot L^{-1}$ for the phenylhydrazine solution and 1.986 ($\pm$ 0.002) $mol \cdot L^{-1}$ for the benzophenone solution. Uncertainties were obtained from propagating errors due to sample preparation (weighing, filling and pipetting) [21]. Glacial acetic acid (Carlo Erba Reagents for analysis, certified 100.0%) was directly used from its original bottle (17.483 $mol \cdot L^{-1}$) and no sampling error needed to be propagated.

5.4.3 Independently measured pure component spectra

Mid-IR and UV-vis pure component spectra of benzophenone (1.986 $mol \cdot L^{-1}$ for mid-IR, 0.993 $mol \cdot L^{-1}$ for UV-vis), phenylhydrazine (1.468 $mol \cdot L^{-1}$ for mid-IR and UV-vis), acetic acid (2.914 $mol \cdot L^{-1}$ for mid-IR, 8.742 $mol \cdot L^{-1}$ for UV-vis), benzophenone-phenylhydrazone (1.000 $mol \cdot L^{-1}$ for mid-IR, 0.500 $mol \cdot L^{-1}$ for UV-vis) and water (27.704 $mol \cdot L^{-1}$ for mid-IR, 41.556 $mol \cdot L^{-1}$ for UV-vis) were independently measured at 25°C in THF.

As not commercially available, the hydrazone product was synthetised by performing the reaction in acetonitrile (Acros Organics for analysis), in which it is not soluble and crystallises. Two successive re-crystallisations were performed to obtain the pure product. The product purity was assessed by GC to be more than 99% and the structure was confirmed by MS and NMR.

5.4.4 Experimental conditions (a) – dosing *Aa*

The reactor was initially charged with 20 mL of the phenylhydrazine solution and 5 mL of the benzophenone solution, then 5 mL of glacial acetic acid were quickly injected using a syringe. Note that for all experimental conditions (a to e), quick injections were done within 24 seconds. Experimental conditions (a) are similar to those used in previous papers [21,22]. The experiment performed under these conditions (Experiment #1) was analysed assuming batch and semi-batch conditions. Note, when semi-batch conditions are assumed, the first measured spectrum in data matrix **Y** is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, initial concentrations of 0.979 (± 0.007) mol·L^{-1} in phenylhydrazine, 0.331 (± 0.003) mol·L^{-1} in benzophenone and 2.9 (± 0.1) mol·L^{-1} in acetic acid were used, accounting for the dilution due to the fast dosing of *Aa*. Note that this batch mode of analysis, neglecting the conversion of about 0.7% during the fast dosing, was shown to be reasonable (see Table 5.2). For semi-batch conditions, initial concentrations of 1.174 mol·L^{-1} (± 0.004) and 0.397 (± 0.002) mol·L^{-1} in phenylhydrazine and benzophenone respectively, a dosing concentration of 17.483 mol·L^{-1} in acetic acid and a dosing rate of 12.5 (± 0.5) mL·min^{-1} were considered. All experimental conditions are summarised in Table 5.1.

5.4.5 Experimental conditions (b) – dosing *B* (1.986 mol·L^{-1})

In these experimental conditions, 5 mL of the benzophenone solution were quickly dosed into the reactor, which was initially filled with 20 mL of the phenylhydrazine solution and 5 mL of glacial acetic acid. Three experiments (Experiments #2 to #4) were carried out under these conditions (see Table 5.1) and were analysed assuming batch conditions, using initial concentrations of 0.979 (± 0.007) mol·L^{-1} in phenylhydrazine, 0.33 (± 0.01) mol·L^{-1} in benzophenone and 2.91 (± 0.02) mol·L^{-1} in acetic acid. Experiment #2 was also analysed assuming semi-batch conditions, with initial concentrations of 1.174 (± 0.004) mol·L^{-1} in phenylhydrazine, 3.50 (± 0.02) mol·L^{-1} in acetic acid, a dosing concentration of 1.986 (± 0.002) mol·L^{-1} in benzophenone and a dosing rate of 12.5 (± 0.5) mL·min^{-1}.

Table 5.1: Experimental conditions (a) to (e), and related methods of analysis (batch, semi-batch analysis).

Experimental conditions [1]	Dosed volumes [mL]			Initial volumes [mL]			Data analysis [2]	Dosing concentrations [mol·L^{-1}] [3]			Initial concentrations [mol·L^{-1}] [3]			# Exp.
	B	*P*	*Aa*	*B*	*P*	*Aa*		*B*	*P*	*Aa*	*B*	*P*	*Aa*	
(a)	-	-	5 [4]	5	20	-	Semi-batch	-	-	17.483 [5]	0.397 (± 0.002)	1.174 (± 0.004)	-	1
							Batch	-	-	-	0.331 (± 0.003)	0.979 (± 0.007)	2.9 (± 0.1)	1
(b)	5 [4]	-	-	-	20	5	Semi-batch	1.986 (± 0.002)	-	-	-	1.174 (± 0.004)	3.50 (± 0.02)	2
							Batch	-	-	-	0.33 (± 0.01)	0.979 (± 0.007)	2.91 (± 0.02)	2 – 4
(c)	5 [4, 6]	-	5 [4, 6]	-	20	-	Semi-batch	1.986 (± 0.002)	-	17.483 [5]	-	1.468 (± 0.001)	-	5
(d)	-	20 [7]	-	5	-	5	Semi-batch	-	1.468 (± 0.001)	-	0.993 (± 0.006)	-	8.74 (± 0.05)	6
							Batch	-	-	-	0.331 (± 0.009)	0.98 (± 0.05)	2.91 (± 0.08)	6
(e)	3 [4, 8]	-	-	-	20	5	Batch	-	-	-	0.20 (± 0.01)	0.98 (± 0.01)	2.91 (± 0.03)	7

[1] Experimental conditions are described in the following sections: (a) in Section 5.4.4, (b) in Section 5.4.5, (c) in Section 5.4.6, (d) in Section 5.4.7, (e) in Section 5.4.8

[2] All experiments were performed under semi-batch conditions (fast dosing) but analysed assuming either batch or semi-batch conditions. When semi-batch conditions are assumed, the first measured spectrum in **Y** is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, concentrations are given accounting for the dilution due to dosing

[3] Concentrations are rounded to their last significant digit
Uncertainties are calculated from propagating errors due to sample preparation (weighing, filling and pipetting) [21]

[4] Volume added at a dosing rate of 12.5 (± 0.5) mL·min^{-1}

[5] No sampling error was propagated for *Aa* as it was used from its original bottle

[6] Species *B* and then *Aa* were dosed sequentially

[7] Volume added at a dosing rate of 50 (± 2) mL·min^{-1}

[8] A mixture of 3 mL of the benzophenone stock solution and 2 mL THF was dosed

5.4.6 Experimental conditions (c) – sequential dosing of *B* and *Aa*

The reactor was initially filled with 20 mL of the phenylhydrazine solution. A first quick dosing of 5 mL of benzophenone was performed. Subsequently, 5 mL of glacial acetic acid were quickly dosed. One experiment (Experiment #5) was carried out under these conditions and analysed assuming semi-batch conditions (see Table 5.1), with initial concentrations of 1.468 mol·L^{-1} (± 0.001) in phenylhydrazine, dosing concentrations of 1.986 (± 0.002) mol·L^{-1} and 17.483 mol·L^{-1} in benzophenone and acetic acid respectively, added at dosing rates of 12.5 (± 0.5) mL·min^{-1}.

5.4.7 Experimental conditions (d) – dosing *P*

Here, 5 mL of the benzophenone solution were initially mixed in the reactor with 5 mL glacial acetic acid. Then, 20 mL of the phenylhydrazine solution were quickly dosed. The experiment performed under these conditions (Experiment #6) was analysed assuming batch and semi-batch conditions (see Table 5.1). For batch conditions, initial concentrations of 0.98 (± 0.05) molL^{-1} in phenylhydrazine, 0.331 (± 0.009) mol·L^{-1} in benzophenone and 2.91 (± 0.08) mol·L^{-1} in acetic acid were used. For semi-batch conditions, initial concentrations of 0.993 (± 0.006) mol·L^{-1} in benzophenone, 8.74 (± 0.05) mol·L^{-1} in acetic acid, a dosing concentration of 1.468 (± 0.001) mol·L^{-1} in phenylhydrazine and a dosing rate of 50 (± 2) mL·min^{-1} were considered.

5.4.8 Experimental conditions (e) – dosing *B* (1.192 mol·L^{-1})

In this fifth experimental procedure, 20 mL of the phenylhydrazine solution and 5 mL of glacial acetic acid were initially charged in the reactor and a mixture of 3 mL of the benzophenone solution and additional 2 mL THF was quickly dosed. This was done in order to maintain an approximately constant THF concentration to avoid baseline changes due to the change of the reference spectrum (THF). The experiment performed under these conditions (Experiment #7) was analysed under batch conditions (seeTable 5.1), with initial concentrations of 0.98 (± 0.01) mol·L^{-1} in phenylhydrazine, 0.20 (± 0.01) mol·L^{-1} in benzophenone and 2.91 (± 0.03) mol·L^{-1} in acetic acid. These conditions resemble conditions (b) except that the dosing concentration of *B* represents only 3/5 compared to the one used in experimental conditions (b), i.e. 1.192 mol·L^{-1}.

5.5 Results and discussion

As the reaction studied in this paper only involves one single kinetic step, several matrices and vectors collapse to vectors and scalars respectively. However, for brevity, the boldface capital and boldface lowercase notations for matrices and vectors respectively are maintained, e.g. vector **k** is comprised by only one element k_1. The few Matlab lines required for the calculation of most equations of the following sections are given in Section 5.7 (Appendix).

5.5.1 Analytical solution for Ω

The reaction of benzophenone with phenylhydrazine catalysed by acetic acid, as described in Figure 5.1, includes *ns* = 5 species (*B*, *P*, *Aa*, *BP* and H_2O) involved in *nr* = 1 reaction. The reaction follows a third order rate law **Φ** (*nt* x 1), with partial orders of one for each reactant (*B*, *P*, *Aa*) [21,22]. As water is transparent in the followed range of mid-IR (see Figure 5.2a) and UV-vis (see Figure 5.2b), it can be omitted and considered as non-absorbing, formally including for this species a known spectrum comprised by zeros into the analysis (Strategy 2). Thus, the number of species is reduced to (*ns*–*nks*) = (5–1) = 4. For simplicity, this reduced number of species (*ns*–*nks*) is re-defined as *ns*, i.e. we assume a kinetic model with *ns* = 4 species (*B*, *P*, *Aa*, *BP*). For a description of the differential equations describing the concentration profiles, we refer to [21] (Chapter 3).

$$B + P + Aa \xrightarrow{k_1} BP + Aa \quad \text{with} \quad \phi_t = k_1 c_{t,B} c_{t,P} c_{t,Aa} \tag{5.6}$$

Matrices **E** (1 x 4) and **N** (1 x 4), as given in Equations 5.7 and 5.8, describe the kinetic hard-model of Equation 5.6 in terms of reactant and stoichiometric coefficients respectively.

$$\mathbf{E} = \begin{bmatrix} \overset{B}{1} & \overset{P}{1} & \overset{Aa}{1} & \overset{BP}{0} \end{bmatrix} \tag{5.7}$$

$$\mathbf{N} = \begin{bmatrix} \overset{B}{-1} & \overset{P}{-1} & \overset{Aa}{0} & \overset{BP}{1} \end{bmatrix} \tag{5.8}$$

The initial concentrations $\mathbf{c_0}$ can be defined symbolically in order to account for all different initial conditions used in Sections 5.4.4 to 5.4.8.

$$\mathbf{c_0} = \begin{bmatrix} \overset{B}{c_{0,B}} & \overset{P}{c_{0,P}} & \overset{Aa}{c_{0,Aa}} & \overset{BP}{0} \end{bmatrix} \tag{5.9}$$

Using matrices **E** and **N**, matrix $\mathbf{\Omega}$ (5 x 4), as introduced in Equation 5.2, can be calculated as a function of k_1, $\mathbf{c_0}$ and μ.

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\cdot \mathbf{E}^T} k_1 \mathbf{N}}{\mathbf{c_0}} \end{bmatrix} = \begin{bmatrix} \overset{B}{-\mu k_1} & \overset{P}{-\mu k_1} & \overset{Aa}{0} & \overset{BP}{\mu k_1} \\ -\mu k_1 & -\mu k_1 & 0 & \mu k_1 \\ -\mu k_1 & -\mu k_1 & 0 & \mu k_1 \\ \hline -k_1 & -k_1 & 0 & k_1 \\ \hline c_{0,B} & c_{0,P} & c_{0,Aa} & 0 \end{bmatrix} \tag{5.10}$$

As can be easily seen in Equation 5.10, row five of $\mathbf{\Omega}$ is linearly independent from the others, and either of rows one to three are linearly dependent on row four. Consequently, $\mathbf{\Omega_{ind}}$ can, for example, be comprised by rows four and five, for which $\ker \mathbf{\Omega}_{\mathbf{ind}}^{T} = \mathbf{0}$.

$$\mathbf{\Omega_{ind}} = \begin{bmatrix} \overset{B}{-k_1} & \overset{P}{-k_1} & \overset{Aa}{0} & \overset{BP}{k_1} \\ c_{0,B} & c_{0,P} & c_{0,Aa} & 0 \end{bmatrix} \tag{5.11}$$

The linear dependencies in **C** can be predicted from ker $\mathbf{\Omega}$ (4 x 2), which does neither depend on **k** nor on μ. The matlab lines for Equations 5.7 – 5.12, making use of the symbolic toolbox [30], are given in Section 5.7.1 (Appendix).

$$\ker \mathbf{\Omega} = \begin{bmatrix} -\alpha & -\beta \\ 1 & 0 \\ 0 & 1 \\ 1-\alpha & -\beta \end{bmatrix} \begin{matrix} B \\ P \\ Aa \\ BP \end{matrix} \qquad \text{with } \alpha = \frac{c_{0,P}}{c_{0,B}} \text{ and } \beta = \frac{c_{0,Aa}}{c_{0,B}} \tag{5.12}$$

The dimension of the kernel indicates that the matrix of concentration profiles **C** for this mechanism is two times rank-deficient. The absence of rows comprised by zeros only in ker $\mathbf{\Omega}$ (and in ker **C**) leads to the conclusion that all reactive species (*B*, *P*, *Aa* and *BP*) are involved in the linear dependencies of **C**.

Two mass balances for this mechanism can be calculated according to the definition of the kernel, i.e. **C** (ker **C**) = **0**.

$$(-\alpha)c_{t,B} + (1)c_{t,P} + (0)c_{t,Aa} + (1-\alpha)c_{t,BP} = 0 \quad (5.13)$$

$$(-\beta)c_{t,B} + (0)c_{t,P} + (1)c_{t,Aa} + (-\beta)c_{t,BP} = 0 \quad (5.14)$$

Coefficients in parenthesis, representing the values given by the kernel of **Ω** (see Equation 5.12), can alternatively be calculated by mass balance equations, as described in Section 5.7.6 (Appendix).

To treat this rank deficiency problem in **C** and to fit pure component spectra, one of the four strategies described in Sections 5.3.1 – 5.3.4 has to be applied. For reasons that will become obvious in Section 5.5.2, no pure strategies (2), (3) and (4) were applied but they were combined with Strategy (1).

5.5.2 Independently measured pure component spectra

Pure component spectra of species *B*, *P*, *BP* and H_2O, independently measured as outlined in Section 5.4.3, are shown in Figure 5.2a (mid-IR) and Figure 5.2b (UV-vis). As molar absorptivities strongly depend on the solvent and can be affected by baseline shifts, deviations have to be expected between these pure spectra (measured in plain THF as reference) and those measured in a mixture of solvents comprised by acetic acid and THF in a molar ratio of approximately 1:4, as described in the experimental procedures (a) to (e).

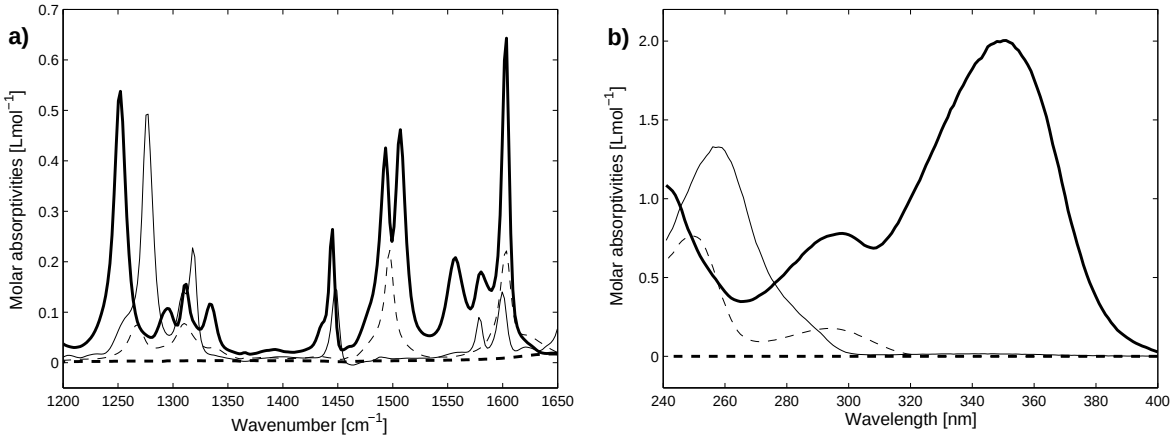

Figure 5.2: Independently measured pure component spectra of benzophenone *B* (——), phenylhydrazine *P* (- - -), benzophenone-phenylhydrazone *BP* (——) and water (- - -) in mid-IR (a) and in UV-vis (b) with THF as background.

For all experimental conditions described in Section 5.4, dosing *Aα*, *B* or *P* affects significantly the molar ratio of THF and acetic acid, i.e. the medium properties, and thus may also alter the pure component spectra.

To counteract this solvent effect, a preliminary experiment (Experiment #1) was performed under experimental conditions (a). This experiment was analysed assuming semi-batch conditions using strategies (1), setting species *B* and *P* uncoloured, and Strategy (3), dosing *Aα*, and optimising the rate constant **k**. As the dosing is fast and only occurs between the acquisition of two reaction spectra, solvent and baseline effects are expected to translate into the resolved fitted component spectrum of acetic acid, as its concentration stays constant with respect to THF after dosing (catalytic effect). This is shown in Figure 5.3a (mid-IR) and Figure 5.3b (UV-vis). Note that, despite this solvent effect, the fitted mid-IR spectrum of *Aα* still shows the distinct pattern of the independently measured one, and the fitted UV-vis spectrum of *Aα* just obtains some small contributions compared to the non-absorbing independently measured one.

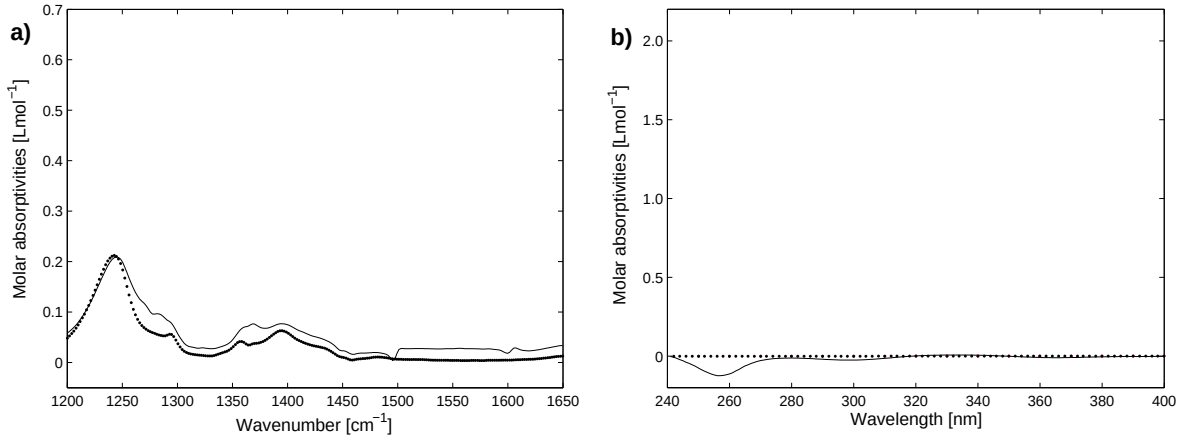

Figure 5.3: Fitted (—) and independently measured (•) mid-IR (a) and UV-vis (b) component spectrum of acetic acid (*Aα*) from the analysis of Experiment #1 assuming semi-batch conditions (*Aα* dosed) and using Strategy (1), setting species *B* and *P* uncoloured.

The fitted pure spectrum of *Aα* (in mid-IR and UV-vis) obtained through the analysis of this preliminary experiment can subsequently be used, together with the independently measured pure spectra of *B*, *P* and *BP*, to reconstitute the fitted component spectra for the experiments of Sections 5.5.3, 5.5.4 and 5.5.6 obtained under different experimental conditions (b and e) and using different strategies (1, 2 and 4). Thus, in first approximation, the deviations due to the change of the solvent mixture of THF and acetic acid do not have to be modelled more deeply.

The optimised rate constants (in mid-IR and UV-vis) obtained for this preliminary experiment analysed under semi-batch conditions are close to literature values as shown in Table 5.2. Differences in rate constants compared to previous works [21,22] are likely due to differences in the purity of the chemicals. This was confirmed by a limited investigation showing that the fitted rate constants depend on the batch to batch variability in purity of the purchased THF and phenylhydrazine.

For the sake of completeness, Experiment #1 was also analysed assuming batch conditions and using Strategy (1), treating species *P* and *Aα* as uncoloured. This analysis resulted in optimised rate constants that are very similar compared to the values obtained when the experiment was analysed assuming semi-batch conditions (see Table 5.2). This confirmed the applicability of a batch analysis for this fast semi-batch experiment.

5.5.3 Application of Strategy (1) – defining uncoloured species

To analyse the reaction of benzophenone with phenylhydrazine by Strategy (1), experimental conditions (b) were used, i.e. species *B* was quickly dosed into a mixture of *P* and *Aα*. Three experiments (Experiments #2 – #4) were performed and analysed individually assuming batch conditions. If the rate constants are not a priori known, $\mathbf{\Omega}$ (5 x 4) and ker $\mathbf{\Omega}$ (4 x 2) can be calculated as a function of $\mathbf{k}$, $\mathbf{c_0}$ and μ. This leads to the general solution given in Equations 5.10 and 5.12.

Alternatively, $\mathbf{\Omega}$ and ker $\mathbf{\Omega}$ can be calculated by defining $\mathbf{k} = \mathbf{1}$, as any strictly positive value does not affect rows of zero in the kernel of the augmented matrix [17]. Using, for example, μ = 2 and the initial conditions described in Section 5.4.5, i.e. $\mathbf{c_0}$ = [0.33 0.979 2.91 0], $\mathbf{\Omega}$ can be calculated numerically.

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\cdot \mathbf{E}^{\mathrm{T}}}\ \mathbf{1N}}{\mathbf{c_0}} \end{bmatrix} = \left[\begin{array}{cccc} \multicolumn{4}{c}{\begin{array}{cccc} B & P & A\alpha & BP \end{array}} \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -1 & -1 & 0 & 1 \\ \hline 0.33 & 0.979 & 2.91 & 0 \end{array}\right] \tag{5.15}$$

Table 5.2: Optimised rate constants for different experimental conditions (a to e) and Strategies (1 to 4).

# Exp.	Experimental conditions [1]	Data analysis [2]	Dosed species [2]	Applied Strategy [3]	mid-IR		UV-vis	
					k_1 [4, 5]	Published k_1 [4]	k_1 [4, 5]	Published k_1 [4]
1	(a)	Semi-batch	*Aa*	(1 & 3)	1.60 (± 0.04)	1.74 (± 0.05) [10]	1.65 (± 0.04)	1.77 (± 0.03) [10]
1	(a)	Batch	*Aa*	(1)	1.59 (± 0.07)	1.40 [11]	1.65 (± 0.07)	1.51 [11]
2	(b)	Semi-batch	*B* [6]	(1 & 3)	1.60 (± 0.03)		1.64 (± 0.03)	
2	(b)	Batch	*B* [6]	(1)	1.58 (± 0.02)		1.63 (± 0.02)	
3	(b)	Batch	*B* [6]	(1)	1.61 (± 0.02)		1.67 (± 0.02)	
4	(b)	Batch	*B* [6]	(1)	1.58 (± 0.02)		1.62 (± 0.02)	
2	(b)	Batch	*B* [6]	(1 & 2)	1.58 (± 0.02) [8]		1.63 (± 0.02) [8]	
5	(c)	Semi-batch	*B* [6] + *Aa*	(3)	1.59 (± 0.06)		1.64 (± 0.06)	
6	(d)	Semi-batch	*P*	(1 & 3)	1.62 (± 0.06)		1.67 (± 0.06)	
6	(d)	Batch	*P*	(1)	1.61 (± 0.09)		1.66 (± 0.09)	
7	(e)	Batch	*B* [7]	(1)	1.55 (± 0.02)		1.57 (± 0.02)	
2 & 7	(b & e)	Batch	*B* [6,7]	(1 & 4)	1.57 (± 0.02) [9]		1.62 (± 0.02) [9]	

[1] Experimental conditions are described in the following sections: (a) in Section 5.4.4, (b) in Section 5.4.5, (c) in Section 5.4.6, (d) in Section 5.4.7, (e) in Section 5.4.8

[2] All experiments were performed under semi-batch conditions (fast dosing) but analysed assuming either batch or semi-batch conditions. When semi-batch conditions are assumed, the first measured spectrum in **Y** is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, concentrations are given accounting for the dilution due to dosing

[3] Strategies are described in the following sections: (1) in Section 5.3.1 (defining uncoloured species), (2) in Section 5.3.2 (including known pure spectra), (3) in Section 5.3.3 (dosing), (4) in Section 5.3.4 (second order global analysis)

[4] in $L^2 \cdot mol^{-2} \cdot s^{-1} \times 10^{-4}$

[5] Uncertainties are calculated according to the propagation of errors developed in Reference [21]

[6] Dosing concentration in *B* of 1.986 $mol \cdot L^{-1}$

[7] Dosing concentration in *B* of 1.192 $mol \cdot L^{-1}$ obtained by dosing a mixture of 3 mL of the benzophenone stock solution and 2 mL THF

[8] Uncertainties in the provided pure spectrum are not propagated

[9] Uncertainty based on the residuals and on the errors in the initial concentrations of Experiment #7

[10] Reference [21]

[11] Reference [22]. Note that pseudo-second order rate constants from this reference have been recalculated to third order rate constants by division with the excess catalyst concentration

Now, the linear dependencies in **C** can be predicted from the kernel of $\boldsymbol{\Omega}$.

$$\ker \boldsymbol{\Omega} = \begin{bmatrix} 0.8083 & 0.1098 \\ -0.3045 & -0.7356 \\ 0.0108 & 0.2350 \\ 0.5038 & -0.6258 \end{bmatrix} \begin{matrix} B \\ P \\ Aa \\ BP \end{matrix} \tag{5.16}$$

As the dimension of the kernel is two, *ns*–rank($\boldsymbol{\Omega}$) = 4–2 = 2 species have to be defined as uncoloured in Strategy (1). These two species can be freely chosen among the *ns* species, as no row in ker $\boldsymbol{\Omega}$ is comprised by zeros only, i.e. no concentration profile is linearly independent from the others. If, for example, species *Aa* is set uncoloured, its corresponding column is removed from $\boldsymbol{\Omega}$ resulting in Equation 5.17.

$$\boldsymbol{\Omega} = \begin{array}{c} \begin{matrix} B & P & BP \end{matrix} \\ \left[\begin{array}{ccc} -2 & -2 & 2 \\ -2 & -2 & 2 \\ -2 & -2 & 2 \\ -1 & -1 & 1 \\ \hline 0.33 & 0.979 & 0 \end{array}\right] \end{array} \tag{5.17}$$

The kernel of $\boldsymbol{\Omega}$ as written in Equation 5.18 has dimensions (3 x 1).

$$\ker \boldsymbol{\Omega} = \begin{array}{c} \begin{matrix} B & P & BP \end{matrix} \\ \left[\begin{array}{ccc} 0.8024 & -0.2705 & 0.5319 \end{array}\right]^{\mathrm{T}} \end{array} \tag{5.18}$$

The kernel in Equation 5.18 indicates that any of the remaining species (*B*, *P* or *BP*) can be defined as uncoloured. If, for example, species *P* is set uncoloured, its corresponding column is removed from $\boldsymbol{\Omega}$.

$$\boldsymbol{\Omega} = \begin{array}{c} \begin{matrix} B & BP \end{matrix} \\ \left[\begin{array}{cc} -2 & 2 \\ -2 & 2 \\ -2 & 2 \\ -1 & 1 \\ \hline 0.33 & 0 \end{array}\right] \end{array} \tag{5.19}$$

The kernel of $\boldsymbol{\Omega}$ as defined in Equation 5.19 now only contains the trivial solution, i.e. ker **C** = **0**. Thus, Experiments #2 – #4 were analysed defining species *P* and *Aa* as uncoloured.

Individual optimised rate constants are gathered in Table 5.2 and result in mean rate constants of $1.59 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$ (mid-IR) and $1.64 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$ (UV-vis). For comparison, very similar rate constants were obtained for Experiment #2 when analysed assuming semi-batch conditions (defining only species *Aa* as uncoloured since *B* is dosed, i.e. combining Strategies 1 and 3), as shown in Table 5.2.

Spectral validation of the kinetic model for Experiment #2 (analysed assuming batch conditions and using Strategy 1) is done by calculation of $\mathbf{\Omega}$ employing the corresponding optimised rate constants obtained after fitting. As an example, $\mathbf{\Omega}$ is calculated for mid-IR spectroscopy using $k_1 = 1.58 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$ (see Table 5.2).

$$\mathbf{\Omega} = \left[\frac{(\mu \mathbf{1})^{\cdot \mathbf{E}^T} k_1 \mathbf{N}}{\mathbf{c_0}} \right] = \left[\begin{array}{cccc} B & P & Aa & BP \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -1.58 \cdot 10^{-4} & -1.58 \cdot 10^{-4} & 0 & 1.58 \cdot 10^{-4} \\ \hline 0.33 & 0.979 & 2.91 & 0 \end{array} \right] \tag{5.20}$$

Keeping only rows four and five in $\mathbf{\Omega}$ results in matrix $\mathbf{\Omega_{ind}}$ (2 x 4), as described in Equation 5.11 with $k_1 = 1.58 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$, $c_{0,B} = 0.33$ $mol \cdot L^{-1}$, $c_{0,P} = 0.979$ $mol \cdot L^{-1}$ and $c_{0,Aa} = 2.91$ $mol \cdot L^{-1}$. Knowing that species *Aa* and *P* were set uncoloured, matrix $\mathbf{\Omega_{ind}}$ can then be used in Equation 5.3 to calculate matrix $\mathbf{\Delta}$ (2 x 4), i.e. the coefficients for the linear combinations of the (resolved) pure component spectra leading to the (unresolved) fitted mid-IR component spectra of the coloured species ('*B*' and '*BP*').

$$\mathbf{\Delta} = \begin{array}{c} \begin{array}{cccc} B & P & Aa & BP \end{array} \\ \left[\begin{array}{cccc} 1 & 2.967 & 8.818 & 0 \\ 0 & 1.967 & 8.818 & 1 \end{array} \right] \end{array} \begin{array}{l} \text{'}B\text{'} \\ \text{'}BP\text{'} \end{array} \tag{5.21}$$

With the rows '*B*' and '*BP*' indicating the coloured species for which (unresolved) fitted component spectra are obtained, and the columns *B*, *P*, *Aa* and *BP* denoting the absorbing species for which true pure spectra exist. The matlab lines required to set up Equations 5.15 – 5.21 are given in Section 5.7.2 (Appendix).

Equation 5.21 indicates that the fitted component spectrum of the coloured species '*B*' is a linear combination of 1 x pure spectrum of *B*, 2.967 x pure spectrum of *P* and 8.818 x pure spectrum of *Aa*; the fitted component spectrum of the coloured species '*BP*' is a linear combination of 1.967 x pure spectrum of *P*, 8.818 x pure spectrum of *Aa* and 1 x pure spectrum of *BP*. As the rate constants determined for UV-vis and mid-IR as so similar, i.e. $1.63 \cdot 10^{-4}$ compared to $1.58 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$ (Table 5.2), matrices $\mathbf{\Delta}$ (2 x 4) for mid-IR and UV-vis are the same, within the precision of Equation 5.21. Note that *Aa* was allowed to absorb in the UV-vis range to account for baseline and solvent effects, as described earlier.

Figure 5.4a and Figure 5.4b show the good agreement between the fitted component spectra of the coloured species and the predicted linear combinations of the true (resolved) pure spectra for mid-IR and UV-vis.

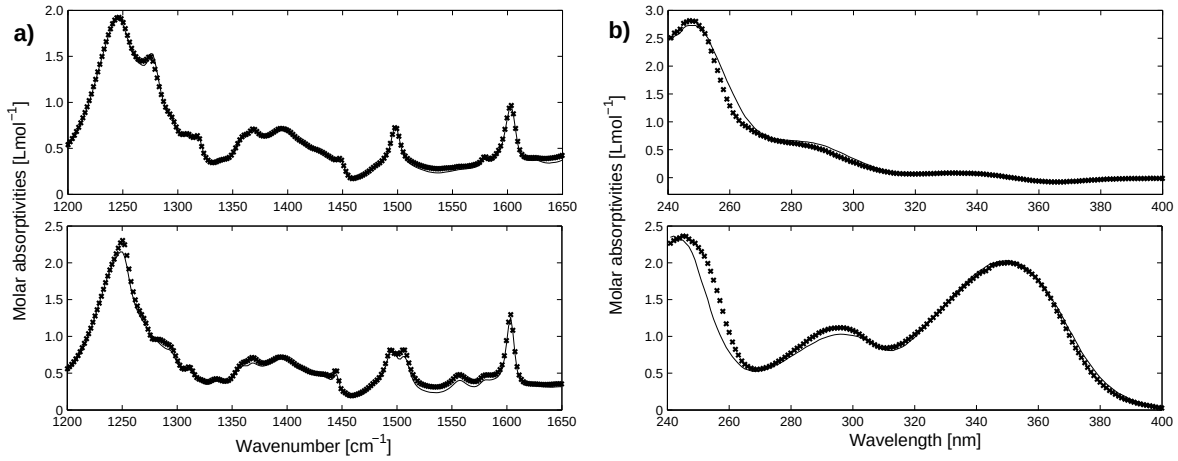

Figure 5.4: Fitted (**—**) and predicted (**x** , Equation 5.21) mid-IR (a) and UV-vis (b) component spectra of coloured species '*B*' (top) and '*BP*' (bottom) from the analysis of Experiment #2 under batch conditions and using Strategy (1), defining species *P* and *Aa* as uncoloured.

5.5.4 Application of Strategy (2) – including known spectra into the analysis

This strategy was also applied to Experiment #2 performed under experimental conditions (b) and analysed assuming batch conditions. If rate constants are not a priori known, $\mathbf{\Omega}$ (5 x 4) and ker $\mathbf{\Omega}$ (4 x 2) can be calculated symbolically (Equations 5.10 and 5.12) or calculated numerically by defining $\mathbf{k} = \mathbf{1}$ (Equations 5.15 and 5.16).

As discussed in Section 5.5.3, two columns need to be removed from matrix **Ω** (or **C**) to enable the calculation of the pseudo-inverse by Equation 5.1. In Strategy (2), these columns could be eliminated by including to the analysis two pure spectra of any of the *ns* species. If however just the pure spectrum of *B* is provided (*nks* = 1), the kernel of **Ω** is not empty, as numerically shown in Equation 5.22 for **k** = **1**.

$$\ker \mathbf{\Omega} = \begin{bmatrix} \overset{P}{0.6879} & \overset{Aa}{-0.2314} & \overset{BP}{0.6879} \end{bmatrix}^{T} \tag{5.22}$$

As ker **Ω** ≠ **0**, instead of including a second known pure spectrum, a second species, e.g. *Aa*, can be defined as uncoloured using Strategy (1) in order to reach ker **Ω** = **0**. So Strategies (1) and (2) are combined. Although full resolution of the calculated spectra cannot be obtained, this puts less constraints and thus less error (due to the known spectra) on the analysis. Thus, Experiment #2 was analysed by providing the pure spectrum of *B* and defining species *Aa* as uncoloured. Fitted rate constants obtained using this Strategy (see Table 5.2) are close to values previously determined in Section 5.5.2 – 5.5.3.

Spectral validation of the kinetic model for Experiment #2 (analysed under batch conditions and using Strategies 1 and 2) is performed by calculating **Ω** (5 x 3) using the corresponding rate constant obtained after fitting. Note that the column dimension of **Ω** is (*ns*–*nks*) = (4–1) = 3, as the pure component spectrum of species *B* was included, and thus this species must now be excluded from Beer's law during the fitting. Keeping only rows four and five of **Ω** results in matrix $\mathbf{\Omega}_{ind}$ (2 x 3), which is shown in Equation 5.23 for UV-vis spectroscopy ($k_1 = 1.63 \cdot 10^{-4}$ $L^2 \cdot mol^{-2} \cdot s^{-1}$ from Table 5.2).

$$\mathbf{\Omega}_{ind} = \begin{bmatrix} \overset{P}{-1.63 \cdot 10^{-4}} & \overset{Aa}{0} & \overset{BP}{1.63 \cdot 10^{-4}} \\ 0.979 & 2.91 & 0 \end{bmatrix} \tag{5.23}$$

Matrix $\mathbf{\Omega}_{ind}$ is then used to calculate matrix **Δ** (2 x 3) and thus the implied linear dependencies of the fitted UV-vis component spectra corresponding to the coloured species '*P*' and '*BP*' (*Aa* set uncoloured). The matlab lines needed to set up Equations 5.22 – 5.24 are given in Section 5.7.3 (Appendix).

$$\mathbf{\Delta} = \begin{matrix} P & Aa & BP \\ \left[\begin{matrix} 1 & 2.972 & 0 \\ 0 & 2.972 & 1 \end{matrix}\right] & \begin{matrix} \text{'}P\text{'} \\ \text{'}BP\text{'} \end{matrix} \end{matrix} \tag{5.24}$$

Due to the similarity between the fitted rate constants in UV-vis and mid-IR, matrices **Δ** for mid-IR and UV-vis are again the same, within the precision of Equation 5.24. Figure 5.5a shows the good agreement between the fitted component spectra and the predicted linear combinations of the true (resolved) pure spectra for mid-IR.

For UV-vis, there are some deviations between the fitted and the predicted component spectra (see Figure 5.5b), particularly in the region of maximum absorbance of species *B*. This indicates that subtracting the contribution of the known spectrum of species *B* from measurement data **Y** is slightly hampered by an imperfection of this known spectrum and does not fully remove the contribution of *B* in **Y**. This was to be expected as the fitted component spectrum of the coloured species '*B*' obtained by Strategy (1) in UV-vis already showed minor discrepancies when compared to the expected linear combinations of the pure spectra of *B*, *P* and *Aa* (see Section 5.5.3 and Figure 5.4b). Naturally, these discrepancies can only increase when an additional constraint is used, such as the inclusion of a known spectrum.

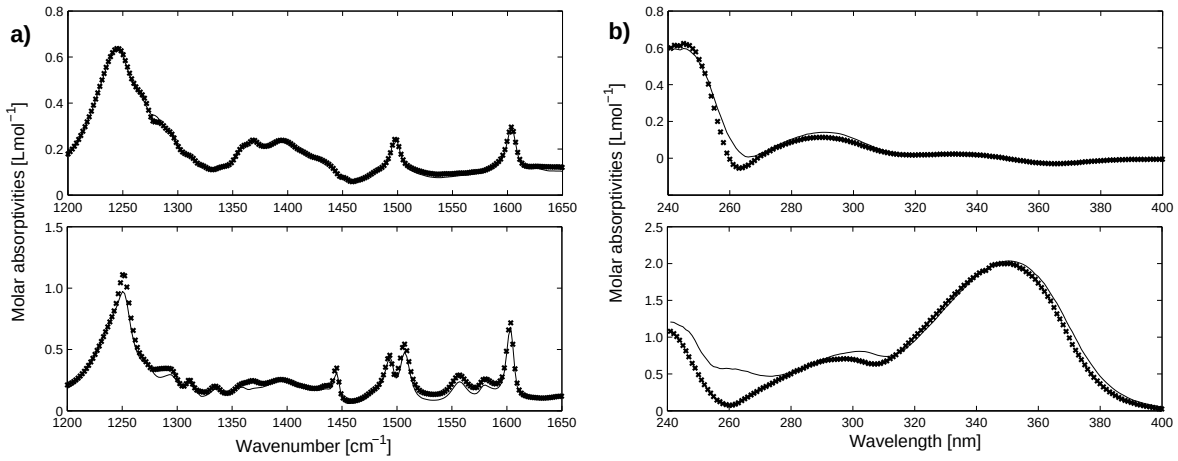

Figure 5.5: Fitted (—) and predicted (**x** , Equation 5.24) mid-IR (a) and UV-vis (b) component spectra of coloured species '*P*' (top) and '*BP*' (bottom) from the analysis of Experiment #2 assuming batch conditions and combining Strategy (2), including the pure component spectrum of species *B*, with Strategy (1), setting species *Aa* uncoloured.

5.5.5 Application of Strategy (3) – dosing one or more species

Experimental conditions (c) were used to present the effects of applying Strategy (3). These experimental conditions involve a first quick dosing of species *B* in *P*, followed by a second dosing, in which species *Aa* is quickly added to the resulting mixture of *B* and *P*. Experiment #5, performed under these conditions, was analysed in semi-batch conditions. With such an experimental design, all species can be included in Beer's law and all fitted pure component spectra are resolved. The dosing conditions are given in Equation 5.25.

$$\mathbf{C}_{\mathbf{in}} = \begin{matrix} B & P & Aa & BP \end{matrix} \\ \begin{bmatrix} 1.986 & 0 & 0 & 0 \\ 0 & 0 & 17.483 & 0 \end{bmatrix} \tag{5.25}$$

Using μ = 2, initial conditions $\mathbf{c_0}$ = [0 1.468 0 0] and defining k_1 symbolically, $\mathbf{\Omega}$ (7 x 4) and ker $\mathbf{\Omega}$ can be calculated according to Equation 5.26. Mass balance equations can alternatively be used to calculate the kernel, as outlined in Section 5.7.7 (Appendix). The matlab lines required for the set up of Equations 5.25 and 5.26 are given in Section 5.7.4 (Appendix).

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\bullet \mathbf{E}^{\mathrm{T}}}\, k_1\mathbf{N}}{\mathbf{c_0}} \\ \hline \mathbf{C}_{\mathbf{in}} \end{bmatrix} = \begin{matrix} B & P & Aa & BP \end{matrix} \\ \left[\begin{array}{cccc} -2k_1 & -2k_1 & 0 & 2k_1 \\ -2k_1 & -2k_1 & 0 & 2k_1 \\ -2k_1 & -2k_1 & 0 & 2k_1 \\ -1k_1 & -1k_1 & 0 & 1k_1 \\ \hline 0 & 1.468 & 0 & 0 \\ \hline 1.986 & 0 & 0 & 0 \\ 0 & 0 & 17.483 & 0 \end{array}\right], \quad \ker \mathbf{\Omega} = \mathbf{0} \tag{5.26}$$

Matrix $\mathbf{\Omega}_{\mathbf{ind}}$ (4 x 4) is comprised by independent rows, such as four, five, six and seven of Equation 5.26 and results in $\ker \mathbf{\Omega}_{\mathbf{ind}}^{\mathrm{T}} = \mathbf{0}$. Calculation of $\mathbf{\Delta}$ (4 x 4) leads to an identity matrix, i.e. all fitted pure component spectra are resolved.

As shown in Table 5.2, fitted rate constants obtained for Experiment #5 (using Strategy 3 and experimental conditions c) are very similar to values previously determined using other strategies. Figure 5.6a (mid-IR) compares the fitted (resolved) pure component spectra with those independently measured in Section 5.5.2. For UV-vis, the fitted pure spectra of the dosed species (*P* and *Aa*) are in very good agreement with their independently measured pure spectra (see Figure 5.6b).

However, for species *B* and *BP* some deviations could be observed between the fitted and the independently measured pure spectra, suggesting that solvent effects and baseline shifts could not be fully corrected.

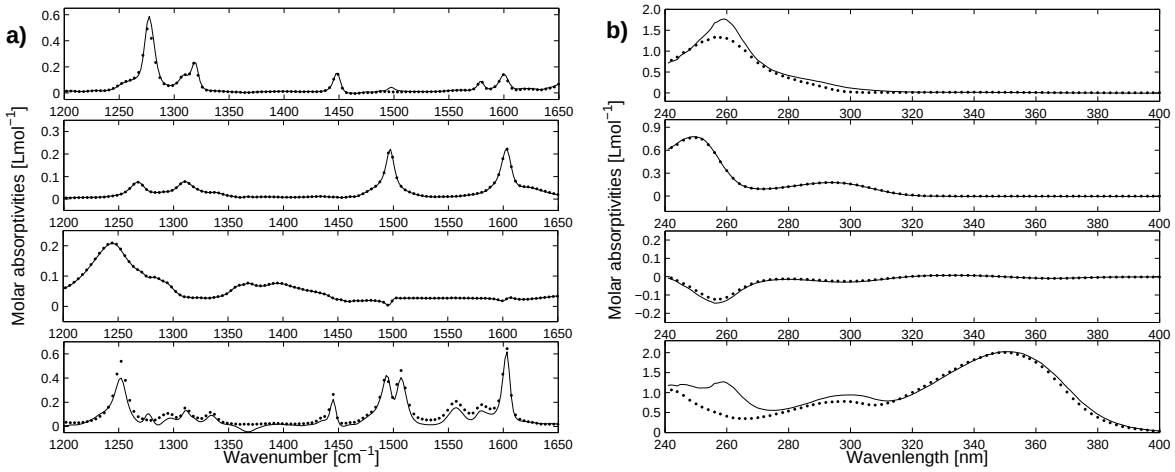

Figure 5.6: Fitted (—) and independently measured (•) mid-IR (a) and UV-vis (b) pure spectra of *B* (1st row), *P* (2nd row), *Aa* (3rd row) and *BP* (4th row) from the analysis of Experiment #5 assuming semi-batch conditions (Strategy 3) and dosing successively species *B* and *Aa* in *P*. Note that the independently measured spectra of *Aa* have been corrected for solvent effects (see Section 5.5.2 and Figure 5.3).

As also shown in Table 5.2, very similar rate constants are obtained for Experiment #6 performed under experimental conditions (d), i.e. when species *P* is dosed into a mixture of *B* and *Aa*, and when this experiment is analysed assuming batch (with *B* and *Aa* uncoloured) and semi-batch (with *Aa* uncoloured) conditions.

5.5.6 Application of Strategy (4) – performing a second order global analysis

For this Strategy, Experiment #2 (Section 5.5.3), performed under experimental conditions (b), was combined with Experiment #7 performed under experimental conditions (e) and just differing from Experiment #2 by its initial concentration in species *B*. Both experiments were analysed individually and also simultaneously by second order global analysis (Strategy 4) assuming batch conditions. For Strategy (4), $\mathbf{\Omega}$ (6 x 4) is calculated using μ = 2, $\mathbf{k}$ = $\mathbf{1}$, initial conditions $\mathbf{c_0}$ = [0.33 0.979 2.91 0] and initial conditions $\mathbf{C_0^{ne}}$ = [0.20 0.98 2.91 0] for the *ne* = 1 additional experiment (Experiment #7).

$$\mathbf{\Omega} = \begin{bmatrix} \dfrac{(\mu\mathbf{1})^{\cdot \mathbf{E}^{\mathrm{T}}}\ \mathbf{1N}}{\mathbf{C_0}} \\ \dfrac{}{\mathbf{C_0^{ne}}} \end{bmatrix} = \begin{array}{c} \begin{matrix} B & P & Aa & BP \end{matrix} \\ \left[\begin{array}{cccc} -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -1 & -1 & 0 & 1 \\ \hline 0.33 & 0.979 & 2.91 & 0 \\ \hline 0.20 & 0.98 & 2.91 & 0 \end{array}\right] \end{array} \tag{5.27}$$

Linear dependencies in **C** are predicted by the kernel of $\mathbf{\Omega}$, which has dimensions (1 x 4).

$$\ker \mathbf{\Omega} = \begin{bmatrix} \overset{B}{0} & \overset{P}{0.6880} & \overset{Aa}{-0.2311} & \overset{BP}{0.6880} \end{bmatrix}^{\mathrm{T}} \tag{5.28}$$

The coefficients in the kernel indicate that the linear dependencies in matrix **C**, i.e. the mass balance (see Equation 5.29), involve the concentrations of species *P*, *Aa* and *BP*. Species *B,* whose initial concentration was varied between the two experiments, is linearly independent from all other species concentrations.

$$\mathbf{C}(\ker \mathbf{C}) = \mathbf{0} \ \Rightarrow\ (0)c_{\mathrm{t},B} + (0.6880)c_{\mathrm{t},P} + (-0.2311)c_{\mathrm{t},Aa} + (0.6880)c_{\mathrm{t},BP} = 0 \tag{5.29}$$

As the rank deficiency is not yet broken, another Strategy, e.g. Strategy (1), has to be applied. According to Equations 5.28 (kernel) and 5.29 (mass balance), any species among *P*, *Aa* or *BP* can be defined as uncoloured. If species *Aa* is treated as uncoloured using Strategy (1), its corresponding column is removed from $\mathbf{\Omega}$ (6 x 4) of Equation 5.27, leading to $\mathbf{\Omega}$ (6 x 3) and ker $\mathbf{\Omega}$ = **0**.

Optimised rate constants obtained for the fitting of Experiment #7 are slightly lower than those received for the fitting of Experiment #2 (see Table 5.2). This explains why the optimised rate constants obtained by Strategy (4) are also lower than the rate constants previously determined with other strategies. Using the rate constant fitted in mid-IR (k_1 = 1.57·10^{-4} L^2·mol^{-2}·s^{-1}), $\mathbf{\Omega}$ is recalculated and reduced to $\mathbf{\Omega_{ind}}$ (3 x 4), by keeping only rows four, five and six of $\mathbf{\Omega}$, which form a vector basis such that $\ker \mathbf{\Omega}_{\mathbf{ind}}^{\mathrm{T}} = \mathbf{0}$.

$$\mathbf{\Omega_{ind}} = \begin{array}{c} \begin{matrix} B & P & Aa & BP \end{matrix} \\ \begin{bmatrix} -1.57\cdot10^{-4} & -1.57\cdot10^{-4} & 0 & 1.57\cdot10^{-4} \\ 0.33 & 0.979 & 2.91 & 0 \\ 0.20 & 0.98 & 2.91 & 0 \end{bmatrix} \end{array} \tag{5.30}$$

The coefficients for the linear combinations of the (resolved) pure component spectra leading to the (unresolved) mid-IR fitted component spectra, **Δ** (3 x 4), are then computed with the constraint that species *Aa* was set uncoloured. The matlab lines needed for the set up of Equations 5.27 – 5.28 and 5.30 – 5.31 are given in Section 5.7.5 (Appendix).

$$\mathbf{\Delta} = \begin{array}{c} \begin{matrix} B & P & Aa & BP \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2.977 & 0 \\ 0 & 0 & 2.977 & 1 \end{bmatrix} \end{array} \begin{matrix} \\ 'B' \\ 'P' \\ 'BP' \end{matrix} \tag{5.31}$$

Equation 5.31 indicates that, under these conditions and the combination of Strategies (1) and (4), the fitted component spectrum of *B* is resolved and the fitted component spectra of coloured species '*P*' and '*BP*' are mixed with the pure spectrum of acetic acid. Figure 5.7a (mid-IR) and Figure 5.7b (UV-vis) compare the fitted component spectra with those independently measured in Section 5.5.2. Component spectra fitted in mid-IR are in good agreement with the predicted linear combinations of the independently measured pure spectra. For UV-vis, some larger deviations can be observed between fitted and predicted component spectra.

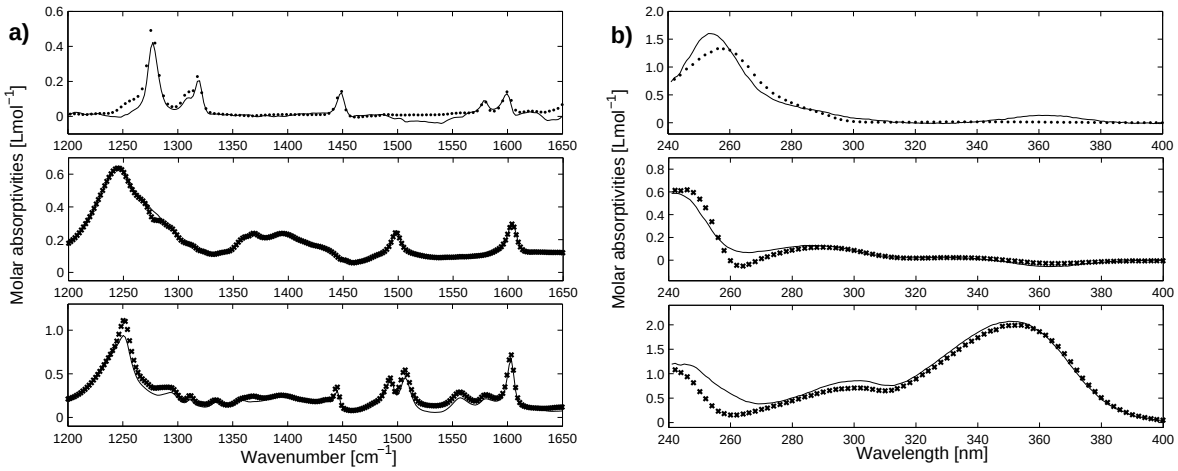

Figure 5.7: Fitted (——), independently measured (•, same as in Figure 5.2) and predicted (**x** , Equation 5.31) mid-IR (a) and UV-vis (b) pure spectra of species *B* (top) and coloured species '*P*' (middle) and '*BP*' (bottom) from the second order analysis of the concatenated Experiments #2 and #7 (assuming batch conditions) measured with different initial concentrations in *B* (Strategy 4) and setting species *Aa* uncoloured (Strategy 1).

5.6 Conclusion

A very intrinsic and important outcome of the direct fitting of spectroscopic data employing a hard kinetic model is the computation of pure component spectra in addition to the kinetic parameters. It is the reproducibility of the fitted kinetic parameters, as well as the correctness of the fitted component spectra, under different experimental conditions that allows the validation of a postulated kinetic mechanism.

However, for kinetic models leading to rank-deficient concentration profiles, the calculation of the component spectra and thus their validation can become very complex. One elegant way to deal with rank deficiency is to define uncoloured species (Strategy 1) as then no or only partial a priori information (known spectra, Strategy 2) or additional experimental design (semi-batch, Strategy 3; variation of initial concentrations, Strategy 4) is required that can potentially perturb the fitting, for example, due to baseline effects and propagated errors. Strategy (1) necessarily leads to fitted component spectra that are complex linear combinations of the true ones.

In the present case study and based on previous theoretical considerations [17], we applied different strategies to allow the kinetic analysis of the rank-deficient acid catalysed reaction of benzophenone with phenylhydrazine in THF. For this condensation reaction, fitted rate constants were highly reproducible, when experimental conditions were varied, e.g. dosing any of the reactants. We have also shown that the coefficients for the linear combinations of the true pure component spectra (matrix $\mathbf{\Delta}$) can be correctly predicted, even when Strategy (1) is combined with any other strategies. This indicates that the kinetics of this reation is well-described by the given mechanism.

Knowing the coefficients for the linear combinations of the true pure component spectra allows a rationalisation of the fitted ones as they can now be assessed on the basis of independently measured pure component spectra (as long as all species can be isolated, as it was possible in this case study).

In many other cases, however, isolation can be a tedious task or sometimes even impossible, particularly for intermediate species. Then, the assessment needs to be done by a peak to peak assignment. For this, it is even more important to know the individual possible contributions of each species to the fitted component spectra.

Acknowledgements

We thank Andreas Dutly from ETH Zurich for the GC-MS measurements and the analysis of chromatograms and MS spectra, and Dr Heinz Rüegger from ETH Zurich for the interpretation of NMR spectra.

5.7 Appendix

5.7.1 Matlab code for Section 5.5.1

```
>> syms c0B c0P c0Aa mu k1;
>> E         = [1 1 1 0];                                    (Equation 5.7)
>> N         = [-1 -1 0 1];                                  (Equation 5.8)
>> one       = ones(size(E'));
>> k         = k1;
>> c0        = [c0A, c0P, c0Aa, 0];                          (Equation 5.9)
>> omega     = [(mu*one).^(E')*diag(k)*N; c0];               (Equation 5.10)
>> omega_ind = omega([4 5], :);                              (Equation 5.11)
>> null(omega);                                              (Equation 5.12)
```

5.7.2 Matlab code for Section 5.5.3

```
>> E         = [1 1 1 0]; N = [-1 -1 0 1]; one = ones(size(E'));
>> k         = 1;
>> mu        = 2;
>> c0        = [0.33, 0.979, 2.91, 0];
>> omega     = [(mu*one).^(E')*diag(k)*N; c0];               (Equation 5.15)
>> null(omega);                                              (Equation 5.16)
>> omega(:, [1 2 4]);                                        (Equation 5.17)
>> null(omega(:, [1 2 4]));                                  (Equation 5.18)
>> omega(:, [1 4]);                                          (Equation 5.19)
>> k         = 1.58e-4;
>> omega     = [(mu*one).^(E')*diag(k)*N; c0];               (Equation 5.20)
>> omega_ind = omega([4 5], :);
>> delta     = inv(omega_ind(:, [1 4]))*omega_ind;           (Equation 5.21)
```

5.7.3 Matlab code for Section 5.5.4

```
>> E         = [1 1 1 0]; N = [-1 -1 0 1]; one = ones(size(E'));
>> k         = 1;
>> mu        = 2;
>> c0        = [0.33, 0.979, 2.91, 0];
>> omega     = [(mu*one).^(E')*diag(k)*N; c0];
>> null(omega(:, [2 3 4]))                                   (Equation 5.22)
>> k         = 1.63e-4;
>> omega     = [(mu*one).^(E')*diag(k)*N; c0];
>> omega_ind = omega([4 5], [2 3 4]);                        (Equation 5.23)
>> delta     = inv(omega_ind(:, [1 3]))*omega_ind;           (Equation 5.24)
```

5.7.4 Matlab code for Section 5.5.5

```
>> syms k1;
>> E         = [1 1 1 0]; N = [-1 -1 0 1]; one = ones(size(E'));
>> k         = k1;
>> mu        = 2;
>> c0        = [0 1.468 0 0];
>> Cin       = [1.986 0 0 0; 0 0 17.483 0];                  (Equation 5.25)
>> omega     = [(mu*one).^(E')*diag(k)*N; c0; Cin];          (Equation 5.26)
```

5.7.5 Matlab code for Section 5.5.6

```
>> E         = [1 1 1 0]; N = [-1 -1 0 1]; one = ones(size(E'));
>> k         = 1;
>> mu        = 2;
>> c0        = [0.331, 0.979, 2.914, 0];
>> C0ne      = [0.199, 0.979, 2.914, 0];
>> omega     = [(mu*one).^(E')*diag(k)*N; c0; C0ne];         (Equation 5.27)
>> null(omega);                                              (Equation 5.28)
>> k         = 1.57e-4;
>> omega     = [(mu*one).^(E')*diag(k)*N; c0; C0ne];
>> omega_ind = omega([4 5 6], :);                            (Equation 5.30)
>> delta     = inv(omega_ind(:, [1 2 4]))*omega_ind;         (Equation 5.31)
```

5.7.6 Calculation of ker $\boldsymbol{\Omega}$ under batch conditions using mass balance equations

Under batch conditions, the mass balance can be written in concentration terms according to the following equation:

$$\mathbf{C_0} = \mathbf{C} - \mathbf{XN} \tag{5.32}$$

The matrix of initial concentrations, $\mathbf{C_0}$ (*nt* x *ns*), is calculated as $\mathbf{1c_0}$, with $\mathbf{c_0}$ (1 x *ns*) being the vector of initial concentrations and **1** a column vector of ones of dimensions (*nt* x 1). Matrix **X** (*nt* x *nr*) denotes the extent of reaction and matrix **N** is defined in Equation 5.8. For details regarding Equation 5.32, we refer to [17] (Chapter 4).

For the chemical system of Equation 5.6, the application of Equation 5.32 leads to the following system:

$$c_{0,B} = c_{t,B} + c_{t,BP} \tag{5.33}$$

$$c_{0,P} = c_{t,P} + c_{t,BP} \tag{5.34}$$

$$c_{0,Aa} = c_{t,Aa} \tag{5.35}$$

Importantly, concentration profiles $c_{t,B}$, $c_{t,P}$, $c_{t,Aa}$ and $c_{t,BP}$ are only linearly dependent if it is possible to write the relationship between these concentrations as homogeneous algebraic equations with time invariant coefficients. In this context, 'homogeneous' indicates that the equations equal the null vector **0**. In such case, the number of these equations defines the dimension of ker $\boldsymbol{\Omega}$ and the time invariant coefficients are the elements of the kernel, such that **C** (ker **C**) = **0**.

For the studied chemical system, a first homogeneous equation with time invariant coefficients (see Equation 5.36) can be calculated according to the ratio between Equations 5.34 and 5.33, denoted (Eq. 5.34 /Eq. 5.33). A second equation (see Equation 5.37) can be obtained by taking the ratio (Eq. 5.35 / Eq. 5.33). Time invariant coefficients of the following equations define the elements of ker $\mathbf{\Omega}$, as previously shown in Equation 5.12.

$$\text{(Eq. 5.34 / Eq. 5.33)}:\ (-\alpha)c_{t,B}+(1)c_{t,P}+(0)c_{t,Aa}+(1-\alpha)c_{t,BP}=0 \quad \text{with } \alpha=\frac{c_{0,P}}{c_{0,B}} \tag{5.36}$$

$$\text{(Eq. 5.35 / Eq. 5.33)}:\ (-\beta)c_{t,B}+(0)c_{t,P}+(1)c_{t,Aa}+(-\beta)c_{t,BP}=0 \quad \text{with } \beta=\frac{c_{0,Aa}}{c_{0,B}} \tag{5.37}$$

These equations are identical to Equations 5.13 and 5.14. As the system of Equations 5.33 – 5.35 is comprised by three equations, two linearly independent ratios can be defined, i.e. the dimension of ker $\mathbf{\Omega}$ is two. For parallelism with Section 5.7.7, the equation obtained from the ratio (Eq. 5.34 / Eq. 5.35) is also calculated.

$$\text{(Eq. 5.34 / Eq. 5.35)}:\ (0)c_{t,B}+(1)c_{t,P}+(-\gamma)c_{t,Aa}+(1)c_{t,BP}=0 \quad \text{with } \gamma=\frac{c_{0,P}}{c_{0,Aa}} \tag{5.38}$$

5.7.7 Calculation of ker $\mathbf{\Omega}$ under semi-batch conditions using mass balance equations

Under semi-batch conditions, the mass balance described in Equation 5.32 requires minor adaptations to take into account the dosed concentrations and the dilution effect. This is shown in the following equation:

$$\mathbf{DC_0}+(\mathbf{I}-\mathbf{D})\mathbf{C_{dos}}=\mathbf{C}-\mathbf{XN} \tag{5.39}$$

Matrix **D** (*nt* x *nt*) is the diagonal matrix of dilution calculated as $v_0\text{DIAG}(\mathbf{v})^{-1}$, with **v** (*nt* x 1) being the vector of volumes and v_0 the initial volume such that $v_0 = v_1$. $\mathbf{C_{dos}}$ (*nt* x *ns*) denotes the matrix of dosing concentrations obtained using $\mathbf{C_{in}}$ (*nf* x *ns*), and **I** (*nt* x *nt*) is the identity matrix. For details regarding Equation 5.39, we refer to [17] (Chapter 4).

When species *B* is dosed alone, Equation 5.39 leads to the following system of equations:

$$d_{t,t}c_{0,B} + \left(1 - d_{t,t}\right)c_{in,B} = c_{t,B} + c_{t,BP} \tag{5.40}$$

$$d_{t,t}c_{0,P} = c_{t,P} + c_{t,BP} \tag{5.41}$$

$$d_{t,t}c_{0,Aa} = c_{t,Aa} \tag{5.42}$$

Here, the ratio (Eq. 5.41 / Eq. 5.40) also leads to a homogeneous equation however with time variant coefficients.

$$\left(\text{Eq. 5.41 / Eq. 5.40}\right):\ \left(-\alpha'\right)c_{t,B} + \left(1\right)c_{t,P} + \left(0\right)c_{t,Aa} + \left(1-\alpha'\right)c_{t,BP} = 0$$

$$\text{with } \alpha' = \frac{d_{t,t}c_{0,P}}{d_{t,t}c_{0,B} + \left(1 - d_{t,t}\right)c_{in,B}} \tag{5.43}$$

Coefficients obtained for the ratio (Eq. 5.41 / Eq. 5.42) are still time invariant and the same as previously defined in Equation 5.38 for batch conditions.

$$\left(\text{Eq. 5.41 / Eq. 5.42}\right):\ \left(0\right)c_{t,B} + \left(1\right)c_{t,P} + \left(-\gamma\right)c_{t,Aa} + \left(1\right)c_{t,BP} = 0 \qquad \text{with } \gamma = \frac{c_{0,P}}{c_{0,Aa}} \tag{5.44}$$

As there is only one remaining equation left that can be written as a function of time invariant coefficients, $\ker \mathbf{\Omega}$ is reduced to dimension one, with its elements given in Equation 5.44.

When species *B* and *Aa* are dosed sequentially (as described in Section 5.5.5), Equation 5.42 has to be modified and replaced by Equation 5.47 in order to take into account the dosed concentration of *Aa*.

$$d_{t,t}c_{0,B} + \left(1 - d_{t,t}\right)c_{in,B} = c_{t,B} + c_{t,BP} \tag{5.45}$$

$$d_{t,t}c_{0,P} = c_{t,P} + c_{t,BP} \tag{5.46}$$

$$d_{t,t}c_{0,Aa} + \left(1 - d_{t,t}\right)c_{in,Aa} = c_{t,Aa} \tag{5.47}$$

With this sequential dosing, it is now impossible to set up a homogeneous equation with time invariant coefficients, as any ratio of Equations 5.45, 5.46 and 5.47 leads to time variant coefficients. This is, for example, the case for the ratio (Eq. 5.46 / Eq. 5.47) when compared to Equation 5.44.

$$\left(\text{Eq. 5.46 / Eq. 5.47}\right):\ (0)c_{t,B}+(1)c_{t,P}+\left(-\gamma'\right)c_{t,Aa}+(1)c_{t,BP}=0$$

$$\text{with}\ \ \gamma'=\frac{d_{t,t}c_{0,P}}{d_{t,t}c_{0,Aa}+\left(1-d_{t,t}\right)c_{in,Aa}} \qquad (5.48)$$

As a consequence, the dimension of ker **Ω** is reduced to zero and the kernel is only comprised of the trivial solution (**0**), i.e. linear dependencies in **C** are broken.

5.8 References

[1] P. Geladi, B.R. Kowalski, Anal. Chim. Acta, 185 (1986), 1-17.
[2] M. Maeder, A.D. Zuberbühler, Anal. Chem., 62 (1990), 2220-2224.
[3] E. Sanchez, B.R. Kowalski, J. Chemom., 4 (1990), 29-45.
[4] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, Comput. Chem. Eng., 20 (1996), 1451-1465.
[5] R. Bro, Chemom. Intell. Lab. Syst., 38 (1997), 149-171.
[6] E.R. Malinowski, Factor Analysis in Chemistry, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
[7] M. Maeder, Y.M. Neuhold, Chapter 7 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 218-256.
[8] R. Tauler, A. de Juan, Chapter 11 in Practical Guide to Chemometrics, Gemperline, P. (Ed.), Taylor and Francis, Boca Raton, USA, 2006, 417-474.
[9] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, Trac-Trends Anal. Chem., 26 (2007), 752-765.
[10] M. Maeder, Y.M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, NL, 2007.
[11] E. Furusjö, L.G. Danielsson, Anal. Chim. Acta, 373 (1998), 83-94.
[12] S. Bijlsma, D.J. Louwerse, A.K. Smilde, AIChE J., 44 (1998), 2713-2723.
[13] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst., 59 (2001), 19-31.
[14] G. Puxty, M. Maeder, K. Hungerbühler, Chemom. Intell. Lab. Syst., 81 (2006), 149-164.
[15] V.M. Taavitsainen, H. Haario, J. Chemom., 15 (2001), 215-239.
[16] V.M. Taavitsainen, H. Haario, M. Laine, J. Chemom., 17 (2003), 140-150.
[17] J. Billeter, Y.M. Neuhold, K. Hungerbuhler, Chemom. Intell. Lab. Syst., 95 (2009), 170-187.
[18] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta, 396 (1999), 231-240.
[19] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurenczy, M. Maeder, A.E. Merbach, A.D. Zuberbühler, Anal. Chim. Acta, 298 (1994), 193-201.
[20] R. Dyson, M. Maeder, Y.M. Neuhold, G. Puxty, Anal. Chim. Acta, 490 (2003), 99-108.
[21] J. Billeter, Y.M. Neuhold, L. Simon, G. Puxty, K. Hungerbuhler, Chemom. Intell. Lab. Syst., 93 (2008), 120-131.
[22] A.R. de Carvalho, M.D. Sanchez, J. Wattoom, R.G. Brereton, Talanta, 68 (2006), 1190-1200.
[23] R. Dyson, M. Maeder, G. Puxty, Y.M. Neuhold, Inorg. React. Mech., 5 (2003), 39-46.
[24] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C++ - The art of Scientific Computing, Second Edition, Cambridge University Press, New York, USA, 2005.
[25] K.Q. Levenberg, Appl. Math., 2 (1949), 164-168.
[26] D.W. Marquardt, J. Soc. Ind. Appl. Math., 11 (1963), 431-441.
[27] P.R. Bevington, D.K. Robinson, Data reduction and error analysis for the physical sciences, Third Edition, Mcgraw-Hill, New York, USA, 2003.
[28] D.C. Lay, Linear Algebra and Its Applications, Third Edition, Addison Wesley Higher Education, Boston, USA, 2003.
[29] F. Visentin, S.I. Gianoli, A. Zogg, O.M. Kut, K. Hungerbühler, Org. Proc. Res. Dev., 8 (2004), 725-737.
[30] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007; http://www.mathworks.com

CHAPTER 6

Conclusion and outlook

6.1 Conclusion

Motivated by the needs of chemical and pharmaceutical industry for efficient tools capable to elucidate kinetic models of industrial reactions, two novel chemometric methods for an improved kinetic hard-modelling of multivariate spectroscopic data have been introduced. They had the aims to facilitate the analysis of rank deficient bilinear spectroscopic data, to enable the validation of rank deficient kinetic models, and to improve the determination of uncertainties in the optimised kinetic parameters, e.g. the rate constants.

Uncertainties and error propagation in kinetic hard-modelling

In Chapter 3, a method has been proposed for the rigorous propagation of uncertainties in the experimental conditions (initial concentrations and dosing rates) into the errors of the rate constants. The impact of these experimental uncertainties onto the error of rate constants fitted by implicit direct kinetic hard-modelling has been validated using simulated and experimental data.

Based on simulated data, it has been possible to identify the experimental conditions leading to minimum uncertainties in the optimised rate constants. For example, the rate constant of a second order mechanism can be fitted with a minimum error due to the uncertainties in its initial concentration under exact stoichiometric conditions or when the reactant with the lowest associated uncertainty in its initial concentration is in a reasonable excess (pseudo first order conditions).

The reaction of benzophenone with phenylhydrazine was also repeatedly (17 experiments) investigated by UV-vis and mid-IR spectroscopy under semi-batch conditions, dosing the catalyst (acetic acid). Each measurement was analysed individually and the standard deviation of each fitted rate constant was predicted by the proposed method up to 80% (UV-vis) and 40% (mid-IR) of the experimental standard deviation over all individually fitted rate constants. Also, it was found that the largest contribution to this predicted error resulted from the dosing rate. Possible reasons for the lower error predictability in mid-IR were elaborated.

Systematic prediction of linear dependencies in the concentration profiles and spectral validation of rank deficient mechanisms

A systematic method for the prediction of linear dependencies in the concentration profiles and for the determination of the expected level of spectral resolution has been introduced in Chapter 4. This generalised approach leads to the elucidation of the fundamental relationship between any postulated kinetic model and the corresponding calculated pure component spectra, which has long been a challenge in kinetic hard-modelling of spectroscopic data. In details, the method allows identifying suitable strategies to circumvent or break linear dependencies in the concentration profiles and to optimise rate constants and pure component spectra. Importantly, analytical prediction of linear dependencies allows detecting particular stoichiometric conditions leading to spectral resolution of some species only, as thoroughly discussed in Chapter 4 based on simulated data.

This systematic method is based on the kernel of a time invariant matrix covering potential rank deficiency due to stoichiometry and rate laws, provided that rank deficiency in the spectroscopic data is due to linear dependencies in the concentration profiles and not in the true pure component spectra. This assumption generally holds for mid-IR and mostly also for UV-vis spectroscopy. The method also depends to some extent on the absorbing nature (absorptivities) of the reactive species with respect to the signal to noise ratio.

In Chapter 5, this method was applied to the task of spectral validation of the rank deficient acid catalysed reaction of benzophenone with phenylhydrazine in THF. Various experimental conditions and data strategies were considered, which all led to the successful validation of the postulated kinetic model by the agreement between fitted and independently measured mid-IR and UV-vis pure component spectra. The understanding of the linear combinations of the true pure component spectra allowed a rationalisation of the fitted pure spectra, which could be assessed on the basis of independently measured pure component spectra.

6.2 Outlook

The method of error propagation presented in Chapter 3 has been introduced to propagate uncertainties in the experimental conditions onto the errors of the fitted non-linear parameters. It was applied to Strategies 1 (defining uncoloured species) and 3 (dosing), in Chapters 3 and 5. The method could be extended even to cover effects of Strategy 2 (including known spectra in the analysis), Strategy 4 (performing a second order global analysis) and temperature fluctuations on the uncertainties in the non-linear parameters. As long as provided by the chemical manufacturer, uncertainties in the purity of the initial compounds could also easily be included in the error analysis. Furthermore, uncertainties in the experimental conditions could be propagated into the errors of the linear parameters, e.g. of pure component spectra, to further enhance the spectral assessment of kinetic models.

Experimental conditions minimising the uncertainties in the non-linear parameters have been briefly discussed based on a second order mechanism, in Chapter 3. A generalised method to predict these conditions for any type of kinetic mechanism might be elaborated, ideally on the basis of a time invariant approach.

The effects of uncertainties when applying the novel method to predict linear dependencies in the concentration profiles (Chapters 4 and 5) could be investigated systematically. In particular for Strategy 1 (defining uncoloured species), the effects of uncertainties in initial concentrations on the linear combinations of the true pure component spectra and their implications on the spectral validation of rank deficient mechanisms would be interesting.

Ultimately, incorporation of spectral constraints obtained from quantum chemistry, e.g. prediction of the maximum or zero absorptivity regions using Density Functional Theory (DFT), could prove as valuable independent information to break rank deficiencies and obtain full resolution in the fitted component spectra.

In this dissertation, kinetic hard-modelling of spectroscopic data by modelling the concentration profiles was elaborated assuming ideal conditions of mixing and of temperature control. These ideal conditions were experimentally provided by a small-scale reaction calorimeter (CRC.v4). However, for non-ideal conditions, as encounted in industrial reactors, the modelling of the concentration profiles would certainly require adaptations in order to account for imperfect feeding, stirring and heating control resulting in non-uniform concentrations, variations in temperature and unexpected side reactions. A solution for modelling some of this non-ideal behaviour could be the hard-modelling of the main kinetic reaction and the soft modelling of the side reactions and of the artefacts. An alternative, however computationally more expensive, could be the combination of kinetic hard-modelling with methods of Computational Fluid Dynamics (CFD), which solve first-principles models by discretising the reaction environment in time and space, and by numerically integrating Partial Differential Equations (PDE) on the defined discretised domain.

"L'homme de science le sait bien, lui, que seule la science a pu, au fil des siècles, lui apporter l'horloge pointeuse et le parcmètre automatique sans lesquels il n'est pas de bonheur terrestre possible"

"Man of science knows that, over centuries, it could only be science that brought him the chronometer as well as the parking meter, without which there is no possible happiness on earth"

"Jeder Wissenschaflter weiss, dass es genau die Wissenschaft ist, die ihm über Jahrhunderte sowohl die Wanduhr als auch die Parkuhr brachte, ohne die es keine irdische Beglückung gibt"

Pierre Desproges
French humorist
1939 – 1988

Julien Billeter **CURRICULUM VITAE**

Personal data

Born on June 24, 1978, in Switzerland
Citizen of Männedorf (Zürich)

Education and Degrees

2005 – 2009 **Ph.D. in Sciences** (Doctor of Sciences)
Swiss Federal Institute of Technology Zurich (ETHZ)
Thesis: Chemometric methods for prediction of uncertainties and spectral validation of rank deficient mechanisms in kinetic hard-modelling of spectroscopic data
Advisers: Prof. K. Hungerbühler, Prof. M. Morari, Dr. Y.-M. Neuhold

2003 – 2004 **Master in Business Management** (DEA en gestion d'entreprise)
Business School (HEC), University of Geneva
Thesis: Share buyback programs through allocation of put options in Switzerland
Adviser: Prof. D. Isakov

1998 – 2003 **Master in Chemical Engineering** (Diplôme d'ingénieur chimiste)
Swiss Federal Institute of Technology Lausanne (EPFL)
Thesis: Simulation of a pollution event over Mexico City: February 15, 2002
Advisers: Prof. H. van den Bergh (EPFL), Prof. M. Molina (MIT, USA)

1994 – 1996 **High School Certificate in Latin-English** (Maturité type B)
Ecole Nouvelle de la Suisse Romande, Lausanne

Publications

2009 **Billeter et al., Chemom. Intell. Lab. Syst., 98 (2009), 213 - 226**
Kinetic hard-modelling and spectral validation of rank deficient spectroscopic data: a case study
Billeter et al., Chemom. Intell. Lab. Syst., 95 (2009), 170 - 187
Systematic prediction of linear dependencies in the concentration profiles and implications on the kinetic hard-modelling of spectroscopic data

2008 **Billeter et al., Chemom. Intell. Lab. Syst., 93 (2008), 120 - 131**
Uncertainties and error propagation in kinetic hard-modelling of spectroscopic data

Conferences

2009 **19th Annual Meeting of the Swiss Chemical Society** (SCS)
Lausanne (Switzerland), September 4, oral presentation
11th Scandinavian Symposium on Chemometrics (SSC)
Loen/Stryn (Norway), June 8 - 11, oral presentation

2008 **100th Annual Meeting of the American Institute of Chemical Engineering** (AIChE)
Philadelphia (USA), November 16 - 21, oral presentation
18th Annual Meeting of the Swiss Chemical Society (SCS)
Zürich (Switzerland), September 11, poster presentation
18th International Congress of Chemical and Process Engineering (CHISA)
Prague (Czech Republic), August 24 - 28, oral presentation

2006 **35th Inorganic Reaction Mechanisms Meeting** (IRMM)
Kraków (Poland), January 4 - 7, poster presentation

Diss. ETH No. 18311 (2009)